





TRAIAN ROTARIU

CURS DE  
METODE ȘI TEHNICI DE  
CERCETARE SOCIOLOGICĂ



TRAIAN ROTARIU

C u r s

de

METODE SI TEHNICI DE CERCETARE SOCIOLOGICA



## C U P R I N S

	Pag.
Prefață . . . . .	1
Partea I. PROBLEME GENERALE . . . . .	5
<u>CAPITOLUL I. TERMENII PRINCIPALI AI CERCETĂRII SOCIOLOGICE</u> . . . . .	5
1. Obiectul și specificul sociologiei . . . . .	6
2. Știința sociologică . . . . .	9
3. Domeniul metodologic . . . . .	21
<u>CAPITOLUL II. ETAPELE CERCETĂRII SOCIOLOGICE</u> . . . . .	28
1. Pregătirea cercetării . . . . .	29
2. Valorificarea informației . . . . .	50
<u>CAPITOLUL III. CALITATIV ȘI CANTITATIV ÎN STUDIAREA FENOMENELOR SOCIALE</u> . . . . .	54
1. Precizarea noțiunilor . . . . .	55
2. Scale de evaluare . . . . .	63
3. Aplicațiile matematicii și statisticii în sociologie . . . . .	73
Partea a II-a. METODE DE CULEGERE A INFORMAȚIEI . . . . .	81
<u>CAPITOLUL IV. ANCHETA</u> . . . . .	81
1. Metoda și tehnicile sale . . . . .	82
2. Chestionarul . . . . .	97
3. Erori și surse de erori în anchete . . . . .	110
<u>CAPITOLUL V. Analiza documentelor</u> . . . . .	120
1. Tipuri de documente . . . . .	122
2. Tehnici de analiză . . . . .	132
<u>CAPITOLUL VI. Experimentul</u> . . . . .	144
1. Ce este experimentul . . . . .	146
2. Schema logică a experimentului . . . . .	149
3. Particularitățile experimentării în sociologie și în științele socio-umane, în general . . . . .	154
4. Experimentul de laborator și experimentul de teren . . . . .	162
Partea a III-a. ANALIZA STATISTICĂ A DATELOR . . . . .	166
<u>CAPITOLUL VII. CARACTERISTICI</u> . . . . .	166
1. Distribuții de frecvență . . . . .	167
2. Curbe atașate distribuțiilor de frecvență . . . . .	181







## P R E F A Ț A

Am adunat în acest volum un set de lecții din cadrul cursului de Metode și tehnici de cercetare sociologică, predat studenților din anii I și II și secției de filosofie-istorie, cu speranța că elaborarea unui text scris va contribui la ameliorarea, pe diferite planuri a activității didactice și va facilita pregătirea studenților pentru examen. Necesitatea unui asemenea manual s-a resimțit foarte acut, căci, în ciuda proliferării literaturii sociologice, nu a apărut la noi în țară nici o lucrare al cărei conținut să acopere satisfăcător sfera de probleme tratate în cursul de față.

Mărturisesc de la început că ideea centrală care m-a ghidat în structurarea și selectarea chestiunilor ce vor fi tratate în paginile ce urmează a fost aceea de a oferi viitorilor profesioniști în domeniul științelor sociale un instrumentar care să-i introducă efectiv în universul cercetării sociologice concrete, să-i ajute în înțelegerea corectă a textelor sociologice de această natură și chiar să le permită proiectarea și realizarea unor studii proprii. Din acest motiv, am evitat tratarea unor aspecte filosofice sau istorice ale cercetării sociologice de teren, care se întâlnesc frecvent în lucrări similare, ocupînd, nejustificat, un loc mult prea mare în dauna problemelor metodice propriu-zise. Am considerat că asemenea digresiuni nu-și au locul aici și pentru faptul că studenții noștri urmează un curs amplu de "Materialism istoric", unul de "Doctrină sociologică" și altul de "Epistemologie socială", cursuri care, împreună, oferă o pregătire generală sociologică satisfăcătoare.

Probabil că, sub forma prezentată, textul meu pare încă destul de abstract, presărat cu prea puține exemplificări concrete. Lucrul se explică prin aceea că am reprodus doar materia predată la curs, ea fiind evident îmbogățită și concretizată în seminarii. De altfel, manualul acesta se cere dublat de unul de exerciții și probleme, de un ghid practic al cercetării sociologice empirice, pe care sper să-l pot elabora într-un timp cît mai scurt posibil.



Conținutul cursului este grupat în patru secțiuni principale. Prima cuprinde probleme de natură generală, specificări de termeni și realizează trecerea în revistă a fazelor generale ale unei cercetări. Am acordat aici o atenție aparte raportului dintre calitativ și cantitativ în investigația sociologică (cap.3) pentru motivul că, după părerea mea, multe dispute sociologice pe această temă sînt sterile datorită neînțelegerii de către protagoniști a noțiunilor de bază folosite. Este drept că tratarea mea poate lăsa impresia că privește doar unele aspecte foarte înguste, foarte specifice ale problemei, în timp ce în literatură atributele de calitativ și cantitativ sînt folosite în chestiuni de mai largă anvergură, dar rămîn convins că la baza oricărei distincții între o abordare calitativă și una cantitativă trebuie să se regăsească în ultimă instanță, delimitările simple operate aici.

Partea a doua cuprinde trei capitole despre tot atîtea metode: ancheta, experimentul și analiza documentelor. Textul despre experiment este oarecum diferit de celelalte, în sensul că nu oferă prea multe informații practice, ci, evidențiind esența metodei, pune mai degrabă în lumină dificultățile experimentării în sociologice. Credința mea este că sociologia (dacă nu includem în ea și psihologia socială) rareori are de a face cu situații ce pot fi studiate experimental. Am ținut să subliniez toate exigențele unui adevărat experiment științific pentru că nu o dată cercetătorii din științele sociale își etichetează studiile cu calificativul pretențios de "experimentale", fără o acoperire practică adecvată.

Cu toate că manualele similare apărute în alte țări poartă pe copertă cel mai adesea titlul de "Metode ale științelor sociale" (acoperind deci o arie mai vastă decît cea a sociologiei), reflectînd astfel dificultatea unor delimitări și de metode, nu numai de obicei, între aceste discipline, am evitat o astfel de strategie, în speță o deschidere prea largă spre psihologia socială, nu atît din motive de concepție, cît de spațiu. Așa se explică lipsa unor lecții care să trateze despre interviu, testul sociometric, scale de atitudine, egantioane mici etc., precum și detaliile practice ale experimentării despre care am amintit mai sus. Dacă tot sînt la capitolul omisiuni să mai amintesc că nici metoda observației nu este tratată într-un capitol special, nu pentru faptul că nu



i-aș acorda importanța cuvenită, ci deoarece, cred eu, ea nu posedă un sistem structurat de tehnici și reguli care să justifice un spațiu de expunere metodică prea larg.

Cum se poate ușor constata, ultimele două părți, de natură statistică, dețin o pondere însemnată în economia manualului. Am adoptat un astfel de model pentru mai multe rațiuni. În primul rând sînt convins că fără o abordare de tip statistic a fenomenelor sociale, conținutul de cunoștințe dobîndite prin cercetările empirice nu poate fi decît foarte scăzut. Această idee este deseori susținută în vorbe, dar o analiză atentă a producției noastre sociologice lasă să se întrevadă unele lacune serioase în planul punerii ei în practică. Moștenirea unei pregătiri preponderent filosofice se manifestă și în acest domeniu, unde, cu cîteva excepții, sociologii noștri se mișcă destul de greoi.

În al doilea rînd, cred că noțiunile simple de statistică introduse în curs sînt necesare nu numai unor viitori sociologi, ci oricărei persoane ce ține la atingerea unui nivel de cunoștințe generale cît de cît ridicat. Gîndirea statistică sînt convins că este o componentă esențială a culturii moderne, iar neglijarea ei de către unii intelectuali de formație umanistă nu face decît să-i mențină pe niște coordonate culturale desuete, depășite. Mai mult, mi se pare absolut de neconceput ca viitorii profesori de științe sociale să poată preda satisfăcător elevilor de liceu cunoștințe social-politice și economice fără să cunoască noțiuni elementare de statistică. Din acest punct de vedere, dar și din cel al pregătirii sociologice, mi se pare că volumul de informație statistică al manualului este încă nesatisfăcător. S-ar fi impus, dacă spațiul mi-ar fi permis aceasta, introducerea cel puțin a unor instrumente de analiză a dinamicii fenomenelor sociale.

În al treilea rînd, extensia mai puțin obișnuită a capitolelor statistice într-un curs de "Metode" se explică și prin convingerea mea că omisiunile la alte teme se pot mult mai simplu recupera, de către cei interesați prin lectură individuală. Slabele cunoștințe matematice ale unei părți importante a studenților filosofi îi privează pe aceștia de posibilitatea consultării aproape a oricărui manual de statistică existent la noi. De aceea m-am străduit să expun aceste probleme într-un limbaj cît mai clar, cît mai



puțin formalizat și să explice pe larg semnificația unor noțiuni și formule.

Cursul se încheie cu două capitole ce tratează problematica eșantionării. Deși alegerea eșantionului este o operație ce se efectuează în faza pregătitoare a cercetării, introducerea noțiunilor legate de selecție n-ar fi fost posibilă fără a trece mai întâi prin aspectele statistice din capitolele 7 - 10.

Lista bibliografică prezentată în final conține, în principal lucrări metodologice de bază, cărora li se adaugă tratate de statistică generală și socială, două lucrări generale despre sociologie (Herseni și Szczepanski) și alte câteva lucrări citate în text cu diverse ocazii. Referitor la această listă, încă două precizări. Mai întâi, volumul scrierilor ce ar fi putut fi incluse la bibliografie este foarte mare, dar ținând seama de o serie de considerente am ajuns la concluzia că se impune o drastică selecție a lor. Am ales, așadar, doar lucrări fundamentale apărute în limba română, de preferință cărți, și volume străine ce se găsesc în bibliotecile Clujului. Apoi, în textul cursului am evitat pe cât mi-a stat în putință introducerea de citate și aceasta în scopul de a facilita lectura. Desigur, nu am pretenția că toate ideile ce apar fără trimiteri sînt originale. Dimpotrivă, gradul de originalitate al unui curs, mai ales de profilul celui de față, căruia i se cere să sintetizeze o experiență de cercetare și cunoaștere bine consolidată, nu poate fi, prin definiție, decît foarte modest. Așadar, rostul listei bibliografice de la sfîrșit este dublu: într-un sens, ea indică aria din care au fost sintetizate o parte din ideile lecțiilor (cealaltă parte, mai modestă, provenind din experiența destul de bogată a autorului în cercetarea empirică) iar, în altul, oferă celor interesați posibilitatea lărgirii cunoștințelor prin consultarea unor surse accesibile. Teribilismul citatelor sau al unei bibliografii nesfîrșite nu-și avea rostul aici, cum, de altfel, cred că nu și-l are nici în alte lucrări, care încearcă, în van, prin asemenea metode să-și ridice statutul de științificitate, la care aspiră atîta cercetători din domeniul disciplinelor sociale.



## PARTEA I

### PROBLEME GENERALE

#### Cap. I. TERMENII PRINCIPALI AI CERCETĂRII SOCIOLOGICE

În concepția autorului cursului de față, sociologia poate fi și trebuie inclusă în familia largă a științelor empirice, în sensul că îndeplinește - deși nu în mod perfect și, cu siguranță, în măsură mai mică decât alte științe - cel puțin următoarele patru cerințe fundamentale ce trebuie avute totdeauna în vedere atunci când urmărim să acordăm statutul de știință unei discipline:

(i) Posedă un obiect empiric propriu de cercetare, decupabil în realitate, fie prin evidențierea unor entități sau fenomene specifice, fie prin precizarea unor modalități de abordare, unghiuri de vedere, perspective etc. proprii. Există deci un set de fapte și fenomene ale universului material a căror descriere, explicare și previziune, dintr-o anumită perspectivă, revin acestei științe.

(ii) Posedă un sistem conceptual specific, cu ajutorul căruia este "reflectată" realitatea obiectivă și transformată într-un obiect de cercetare al științei.

(iii) Discursul științific este construit prin respectarea legilor generale ale logicii, astfel încât să se poată testa pe această bază consistența internă a oricărei deducții, teorii etc.

(iv) Există modalități practice prin care se testează validitatea discursului teoretic, adică tehnici de verificare empirică obiectivă a propozițiilor științei.

În cazul special al sociologiei, se cere accentuată importanța fundamentală pe care acest din urmă aspect o prezintă; peste o atare cerință nu se poate trece fără ca disciplina în cauză să se autoexcludă din rîndul științelor care urmăresc să dea seama de o realitate empirică, trecînd în rîndul altor genuri ale meditației umane ca metafizica, teologia etc. Valabilitatea oricărei propoziții științifice trebuie testată, direct sau indirect



prin contactul cu realitatea și, de asemenea, modalitatea respectivă de testare trebuie să aibă un caracter universal, adică să fie realizabilă de orice om de știință cu mijloacele și în condițiile în care a lucrat autorul aserțiunii emise.

Proiectînd ansamblul exigențelor mai sus formulate asupra activității sociologice, ar fi nevoie, pentru a justifica afirmația că ne găsim în fața unei științe, să vedem în ce măsură aceste criterii sînt îndeplinite. Altfel spus, s-ar cere delimitat obiectul sociologiei, examinate structura și funcționalitatea discursului sociologic și analizate componența și eficiența ansamblului de mijloace - domeniul metodologic - prin care sociologia ajunge la descoperirea unor cunoștințe sau efectuează testarea lor empirică.

Asemenea obiectiv nu poate fi desigur realizat în cadrul restrîns al unui curs ca cel de față, care, nici prin profilul său nu urmărește așa ceva, problematica de mai sus constituind obiect de cercetare în mod deosebit pentru filosofia, logica și metodologia științei. În capitolul introductiv prezent voi face doar unele considerații cu caracter foarte general și cîteva precizări terminologice și de concepție, indispensabile pentru lecțiile ce vor urma.

### 1. Obiectul și specificul sociologiei

A preciza într-o singură frază obiectul unei științe - cum deseori au pretenția că o fac autorii de manuale, în deosebi - mi se pare un lucru, în general, dificil (asemenea definiției avînd aproape întotdeauna forma unui cerc vicios), iar pentru sociologie chiar imposibil, dată fiind aria sa problematică extrem de largă și caracterul foarte divers și greu unificabil sub o singură etichetă a aspectelor studiate.

O primă chestiune evidentă este aceea că sociologia este o știință socială, adică, împreună cu alte discipline ca economia, demografia, psihologia socială, științele juridice etc. studiază probleme referitoare la viața indivizilor în societate, activitățile lor, constituirea și "funcționarea" unor instituții sau colectivități ș.a.m.d. Deci spre deosebire de alte științe ale omului, precum psihologia, disciplinele sociale privesc omul ca o ființă socială, interesîndu-se nu atît de trăsăturile individuale

ale ace

specif

Aici p

pot de

larg a

nem, s

oricar

ramuri

spune

ciale

(școal

le, (d

socia

zonă

cele

etc.)

aprec

lă, i

tate

condi

fică

a edu

studi

econ

lor

subsi

din

compl

punct

(stru

siste

re),

te p

cuno





ale acestuia, ci, în special, de raporturile sale cu ceilalți.

O a doua chestiune ar trebui să fie legată de stabilirea specificului sociologiei în familia mai largă a științelor sociale. Aici părerile specialiștilor sînt foarte împărțite și cu greu se pot desprinde cîteva puncte, dacă nu comune tuturor, cel puțin mai larg acceptate. Mai întîi, sub raportul cuprinderii, așa cum spunem, sociologia are evident o sferă mult mai amplă, comparativ cu oricare altă disciplină socială, de unde și evantaiul complex al ramurilor și subramurilor ei. Încercînd o sistematizare, am putea spune că cercetările sociologice pot avea ca obiect: acțiuni sociale (ca munca, educația, propaganda politică etc.), instituții (școala, familia, biserica, partide politice, ș.a.), grupuri sociale, (de la microgrupuri la cele de mari dimensiuni precum clasele sociale), comunitățile umane teritorial determinate (sat, oraș, zonă funcțională etc.) sau o serie de alte fenomene sociale din cele mai diverse (mobilitatea socială, delinvența, sinuciderea etc.). Apoi, în ceea ce privește specificul de abordare, se poate aprecia că sociologia se caracterizează printr-o perspectivă globală, integratoare, raportînd constant obiectele și fenomenele cercetate la întregul social, studiindu-le într-un cadru al cauzelor, condițiilor și efectelor mai amplu, ceea ce face posibilă și justifică apariția unor sociologii de ramură ca cea economică, juridică, a educației ș.a., paralele cu științele sociale specializate în studierea aceluiași fenomen, dar dintr-o perspectivă mai îngustă: economia, științele juridice, pedagogia etc.

Societatea este înțeleasă de marea majoritate a sociologilor ca un sistem complex, alcătuit, la rîndul său, din o serie de subsisteme, care pot integra și ele altele mai mici ș.a.m.d. Una din sarcinile specifice sociologului constă în analiza cît mai completă a unuia sau altuia dintre aceste subsisteme, adică din punctul de vedere al genezei sale, al modului de funcționare (structura internă, raporturile dintre elementele lui și ale sub-sistemului cu altele de pe același palier sau de pe cele superioare), al tendințelor de evoluție.

Întregul complex de unități, fenomene, procese sociale este privit ca avînd o existență obiectivă, independentă de subiectul cunoscător. Sociologia ca știință este posibilă numai în măsura



în care această premisă fundamentală este acceptată, numai dacă viața socială prezintă regularități, legități, numai dacă evenimentele studiate sînt într-un anume fel repetitive, subsumîndu-se unor sisteme cu funcționare relativ constantă, cu alte cuvinte, numai dacă în universul social există o ordine și un determinism immanent.

O asemenea idee nu este întotdeauna și pe deplin acceptată. Anumite orientări filosofice, plecînd de la faptul evident că nici o societate nu există în afara indivizilor umani și că acțiunea din urmă se deosebește fundamental de elementele materiale ce intră în atenția științelor naturii, prin aceea că sînt dotați cu conștiință, sentimente și voință, care le conferă un anume grad de libertate în acțiunile lor individuale (prin alegerea scopurilor, a mijloacelor de acțiune etc.), în comportamentul lor, în aderența sau nonaderența la anumite valori etc., susțin că sociologia, ca și alte discipline socio-umane, s-ar deosebi radical de celelalte științe și că modelul naturalist de știință ar fi inaplicabil aici.

Desigur, nu este locul într-o lucrare ca cea de față nici să expun pe larg asemenea concepții, nici să acord un spațiu prea mare combaterii lor. Pot și trebuie doar să-mi justific opțiunea prin cîteva aserțiuni cu caracter de principiu, care mi se par suficiente în acest context. De fapt, argumentele mele nu au nici o pretenție de originalitate, ele fiind teze de bază ale concepției marxiste despre societate, acceptate astăzi de un număr mare de sociologi de cele mai diverse orientări.

Caracterul obiectiv al fenomenelor și proceselor sociale nu se datorează faptului că acestea s-ar realiza printr-o participare mecanică, automată a oamenilor la ele, că indivizii umani s-ar supune legităților sociale așa cum, de exemplu, pilitura de fier se așează pe liniile de forță ale cîmpului magnetic, ci sînt rezultatul acțiunii indivizilor umani concreți, dotați cu voință, aspirații, trebuințe, etc. Adică, vreau să spun că anumite generalități sînt rezultantele tot mai abunor astfel de mulțimi de acțiuni individuale, aparent divergente sau haotice, fapt din care rezultă și o consecință metodologică importantă pentru științele sociale: în cele mai multe cazuri, metodele și tehnicile de cercetare sînt chemate să producă informații capabile să fie supuse



unei analize de tip statistic, pentru a decela regularitățile de acest gen.

Pe de altă parte, chiar dacă acțiunile umane sînt rezultatul unui act de decizie individual, aceasta nu înseamnă actul respectiv date arbitrar, gratuit. Există scopuri urmărite de majoritatea indivizilor umani, tot așa cum există mijloace mai eficiente pentru atingerea acestor scopuri, ceea ce conferă, în fapt, acțiunilor anumite constante și regularități. Oricît de liber ar părea omul, el este evident împins, cel puțin prin natura sa biologică, să acționeze mai curînd într-o direcție decît în alta.

În afîrșit, dar nu în ultimul rînd, din momentul în care omul și-a început viața socială, nici o comunitate umană nu poate supraviețui dacă indivizii ce-o compun nu sînt "constrînși" la adoptarea anumitor comportamente, valori, norme etc. comune. Prin urmare, orice societate prezintă regularități de funcționare trans-individuale, independente de voința unuia sau altuia din membrii ce-o compun.

Ansamblul acestor trăsături, care ar putea fi cu ușurință lărgit, fie prin detalierea, fie prin completarea principiilor de mai sus, face posibilă abordarea vieții sociale cu mijloace de cercetare științifică, deci existența unor științe sociale. Această posibilitate se transformă în realitate, deci se ajunge efectiv la știință, dacă activitatea cercetătorului se desfășoară în sensul realizării cerințelor subliniate la începutul capitolului."

## 2. Știința sociologică

### 2.1. Structura științei sociologice

Există astăzi un obicei aproape general de a împărți activitatea științifică în două categorii: una teoretică și alta empirică și de a se face apoi de aici o distincție - deși nu întotdeauna prea convingătoare - între nivelul teoretic și cel empiric al enunțurilor, în cadrul tuturor științelor realului.

Cursul de față nefiind unul de filosofie sau logică științei, nu vom între în prea multe detalii, mărginindu-ne la cîteva



constatări generale, suficiente pentru a înțelege problematica cercetării sociologice. Fiecare om are, în linii mari, o imagine relativ clară despre ceea ce este o teorie dintr-o știință oarecare, probabil cea mai familiară și mai bine conturată, din liceu, fiind teoria mecanicii clasice. De asemenea, latura empirică a cercetării a fost atinsă cu ocazia experimentelor și observațiilor în laborator sau pe teren.

În loc de a da o definiție a teoriei, să-l urmărim pe Lazarsfeld (1971) în descrierea pe care o face acestui nivel al științei: "Se începe prin formularea unui număr de concepte fundamentale. Unele dintre ele se pretează la măsurări, altele sînt construcții a priori, a căror validitate rămîne indeterminată la început. Se definesc apoi manipulările la care sînt supuse aceste unități elementare și care vor permite să tragem concluzii noi, pe care le vom verifica la sfîrșit prin observarea faptelor. La un nivel mai elaborat, aceste teorii posedă, în general, două proprietăți suplimentare: pe de o parte, manipulările și concluziile sînt exprimate sub o formă matematică și, pe de alta, demersul tinde să fie 'reductor': proprietățile gazelor se pot explica prin mișcarea moleculelor, proprietățile corpurilor prin structura lor atomică" (p.76).

Validitatea unei teorii, cu excepția celor logico-matematiche, nu rezultă numai din rigoarea logică a construcției sale, ci cum spuneam, mai ales din confruntarea sa cu realitatea, cu datele experienței empirice. Gradul de dezvoltare a unei teorii se măsoară adesea prin apropierea structurii sale interne de cea a unui sistem deductiv sau chiar matematic. Puține sînt însă științele care au reușit să-și construiască un asemenea sistem perfect coerent. Din ignoranță, este destul de răspîndită ideea după care științele naturii în totalitate (sau în marea lor majoritate) ar poseda o astfel de bază teoretică, pe cînd cele socio-umane ar fi lipsite de ea. De aici rezultă firesc ideea că acestea din urmă sînt subdezvoltate, iar pînă la a spune că ele nici n-ar putea ajunge la rigoarea primelor nu este decît un pas. După cunoștințele mele, cazurile fericite de științe ale naturii cu o structură axiomatico-deductivă sînt însă mai degrabă excepții decît regula.



Adevărul este, cred, că foarte multe științe, atât din domeniul fenomenelor naturii cât și din cel al vieții sociale, folosesc mai ales scheme explicative, care nu derivă dintr-o teorie bine pusă la punct. Cel mai adesea, aceste scheme nu pot fi nici perfect coordonate și coroborate, astfel încât, pe baza lor, să se poată ajunge la construcții teoretice de mare generalitate. Desigur, gradul de rigurozitate al discursului nu este același, chiar și în rândul celor mai puțin încheiate. Un astfel de grad ar putea fi evaluat tocmai prin urmărirea măsurii în care schemele teoretice explicative se leagă unele de celelalte: la un pol am avea discursurile axiomatizate, la celălalt cele care produc scheme explicative ad-hoc, pentru fiecare fenomen studiat. Între cele două extreme este un spațiu de joc enorm de mare, unde fiecare știință își găsește un loc pe o poziție mai înaltă sau mai joasă. Atunci când un număr de scheme explicative se "coagulează" în jurul unui nucleu de propoziții de bază, poate lua naștere un subsistem teoretic explicativ al unui fragment de realitate. Se nasc astfel, ceea ce Merton a denumit cu o expresie devenită deja clasică în sociologie, teoriile de rang mediu.

Dacă, referitor la structura științei, distingerea nivelului teoretic de cel empiric nu este o operație prea ușoară (ba chiar există, cum spunem, temeiuri de a crede că nici nu este posibilă o asemenea distincție), în activitatea practică a cercetătorului este mai vizibilă diferența dintre acțiunea de construcție teoretică și cea de investigare empirică a realității.

Urmărind raționamentul lui Lazarsfeld (1971), activitatea teoretică în sociologie poate fi sistematizată prin operațiile următoare:

- stabilirea de scheme clasificatorii precise,
- formularea de concepte complexe care orientează observatorul spre fapte interesante,
- formularea de probleme de cercetare având o mare importanță din punctul de vedere al societății,
- formularea de idei generale asupra manierei în care schimbările actuale se produc sau pot fi provocate,
- previziuni fondate pe descoperiri empirice încă neverificate (ipoteze).



- punerea în relație a faptelor empirice cu altele, ipotetice sau deja verificate (interpretarea)" (pp.76-77).

Activitatea la nivel empiric constă, esențialmente, în abordarea concretă a realității sociale cu ajutorul metodelor și tehnicilor de investigație empirică. Vom considera deci ca fiind de natură empirică operații ca experimentul, anchetele pe bază de chestionar, analiza unor documente a căror informație nu este sociologic sistematizată, prelucrarea unor date statistice primare etc.

Raportul de forțe dintre cele două genuri de activități fundamentale în sociologie, cea teoretică și cea empirică, a fost adesea schimbător. Dacă în secolul trecut și pînă la primul război mondial, se poate afirma că a predominat abordarea teoretică (a fost epoca marilor sisteme sociologice, de regulă speculative), în perioada interbelică, cu deosebire în Statele Unite, a luat un avînd cu totul remarcabil cercetarea empirică. Ar fi aci de subliniat că toate tehnicile de cercetare concretă au fost preluate de sociologii americani din arsenalul europenilor, care le-au utilizat doar sporadic în perioada anterioară și care, cu puțin excepții (printre ele numărîndu-se și școala sociologică de la București a profesorului D.Gusti) au manifestat aceeași atitudine prudentă în cercetarea empirică pînă la al doilea război. Spre sfîrșitul perioadei interbelice, sociologia empirică americană depășise faza de amatorism, astfel că această formă de cercetare a cunoscut rapid o largă răspîndire în toate țările occidentale. Reacțiile împotriva empirismului excesiv n-au întîrziat să apară, mai cu seamă pe bătrînul continent, și așa se face că din anii '60 sociologia caută să găsească justul echilibru între cele două tipuri de activități.

Excesele, atît speculative cît și cele empiriste, au dus astăzi majoritatea sociologilor în situația de a înțelege mai clar:

## 2.2. Raportul dintre teorie și empirie în sociologie

O primă achiziție, aș spune definitivă, a sociologiei contemporane constă în dobîndirea conștiinței sale de a aspira la statutul de știință pozitivă și, prin urmare, în recunoașterea necesității de a se dezvolta concomitent prin consolidarea edificiului



lui său teoretic, prin îmbogățirea și verificarea, forjarea în practică, a arsenalului metodologic și prin acumularea neîncetată de informații asupra realității sociale.

O a doua achiziție a sociologiei ultimelor decenii ar putea fi numită "împăcarea" cercetării empirice cu cea teoretică. Desigur că nici în perioada sociologiei așa-zis speculative și nici în cea preponderant empirică nu a existat un veritabil "divorț" între cele două sfere de activitate, dar este clar că, exceptând anumite cazuri particulare, ce nu fac decât să întărească regula, accentele erau puse prea puternic pe una sau alta din laturi. Actualmente se recunoaște și se subliniază atât importanța teoriei pentru cercetarea concretă, cât și influența de necontestat pe care investigația empirică o are în dezvoltarea teoriei.

În legătură cu primul aspect, aş menționa, în special, ideile sociologului francez Pierre Bourdieu, unul dintre campionii luptei împotriva empirismului plat. În lucrarea Le métier de sociologue (1968), scrisă împreună cu doi colaboratori, din care s-a produs un fragment și în Antologia sociologiei franceze contemporane (1971), după care citez, este spulberată ideea unei cercetări empirice pure, care să nu facă apel la nici un element teoretic, deci neutre față de o teorie sau o concepție despre viața socială. În acest sens se precizează că "orice practică științifică, chiar atunci și mai ales când se revendică orbeste de la empirismul cel mai orb, angajează presupuneri teoretice și că sociologul nu are de ales decât între interogări inconștiente, deci necontrolate și incoerente, și un corp de ipoteze construite metodic în vederea verificării experimentale" (p.192).

Dacă, așa cum spun sociologii francezi, "măsura și instrumentele de măsură și, mai general, toate operațiile practicii sociologice, de la elaborarea chestionarelor și a codificării până la analiza statistică, sînt tot atîtea teorii în act luate ca proceduri de construcție, conștiente sau inconștiente, de fapte și relații între fapte" (p.192), dacă, zic, lucrurile stau în acest fel, atunci devine evident rolul teoriei în toate etapele cercetării empirice, de la alegerea problemei de studiat și până la redactarea raportului final de cercetare. Definirea operațională a noțiunilor, construcția dimensiunilor și a indicatorilor empirici



pentru acestea, numărul și natura întrebărilor dintr-un chestionar, modalitățile de prelucrare a datelor și chiar schema de eșantionare folosită implică o anumită concepție despre realitatea studiată.

Mai cu seamă rezultatele finale ale unei cercetări nu apar niciodată sub formă empirică pură. Toate datele statistice, de pildă, se cer interpretate, și asta înseamnă încă o dată - și în modul cel mai esențial - apelul la o teorie. Celebra butadă după care cu ajutorul statisticii se poate demonstra orice vine să illustreze o stare de fapt; atunci când teoria lipsește din cercetare, aceleași date pot fi uneori prelucrate și prezentate ca să susțină concluzii contradictorii, nu pentru că datele ar fi incoerente, ci pentru că modalitățile de a le prelucra sînt în funcție de concepții diferite despre fenomenele respective.

Concluzia firească ce se impune este că dacă orice cercetare empirică implică astfel de presupoziii teoretice, este preferabil ca ele să fie conștientizate și explicite înaintea travaliului practic. Chiar dacă sociologia nu beneficiază de teorii suficiente de elaborare, este posibilă constituirea, în faza de pregătire a cercetărilor empirice, cel puțin a unui corp de ipoteze, care să ghideze demersul practic și care să ajute la precizarea conceptelor de lucru. Desigur, că, pe de altă parte, o teorie relativ încheată are un rol fundamental pentru cercetările concrete. În cadrul științelor ce posedă practic astfel de teorii, cercetarea empirică este în proporție covârșitoare nu numai subordonată prin concepte acestora, dar și impusă în actele sale concrete, dirijată, canalizată spre problemele ridicate de teorii. În cazul sociologului însă, studiile empirice se efectuează rareori în intenția de testare a unei teorii iar alegerea problemelor se datorează precumpănitor unor factori extrateoretici.

Referitor la rolul cercetării empirice în dezvoltarea teoriilor voi folosi un text tradus în românește al sociologului american R.K.Merton (1972), ale cărui idei acoperă satisfăcător această problemă și care se pot rezuma în felul următor.

După părerea lui Merton, a existat o oarecare unilateralitate în abordarea chestiunii de față, în sensul că discuțiile sociologice au atribuit drept unică funcție majoră cercetării empirice

testa  
odată  
concr  
reali  
verif  
prop  
că nu  
(p.24  
meni  
unde  
lita  
cerc

pita  
unui  
apar  
deo  
fili

uno  
sch  
sch

cer  
Est  
lui  
dic  
și  
inv  
ta  
aba

dis  
mai  
cer  
ces



testarea sau verificarea ipotezelor (p.285). Desigur, nu este niciodată de prisos să se insiste asupra acestei funcții a studiului concret, ea fiind esențială în orice știință. Nici o teorie asupra realității nu se poate justifica fără contactul cu realitatea; verificarea teoriei înseamnă esențialmente testarea empirică a propozițiilor ei. Dar, cum precizează Merton, "cercetarea empirică nu se limitează la rolul pasiv de a verifica și testa teoria" (p.285), și aceasta, îmi permit să adaug, cu atât mai mult în domeniul unei științe în curs de constituire, precum sociologia, unde schemele teoretice suferă, cum am mai spus, de o mare fragilitate. După autorul american, patru ar fi funcțiile majore ale cercetării empirice în dezvoltarea teoriei:

a) Ea inițiază teoria (fenomenul de serendipitate). Serendipitatea este un termen prin care Merton desemnează descoperirea unui fapt neasteptat, aberrant și capital. Neasteptat, pentru că el apare într-o cercetare în care nu se urmărește așa ceva; aberrant, deoarece faptul respectiv este inconsistent cu teoria; capital, fiindcă descoperirea are consecințe importante asupra teoriei.

b) Ea reformulează (reconstruiește) teoria. Cercetarea unor date noi, neglijate pînă atunci, care nu pot fi explicate cu schemele teoretice existente, conduce la elaborarea unor noi scheme explicative.

c) Ea orientează interesul teoretic. Metodele noi de cercetare empirică pot duce la noi direcții de interes teoretic. Este deci înțeleasă aici mai degrabă o influență a instrumentarului de cercetare asupra teoriei. Cred că este surprins în mod judicios faptul că există o interdependență între domeniul teoretic și cel metodologic; nu numai faptele determină teoria, dar și inventarea unor tehnici sau rafinarea altora cunoscute pot orienta interesul teoretic asupra unor domenii neglijate sau, eventual, abandonate speculației pure.

d) Ea ajută la clarificarea conceptelor. Dacă într-un discurs speculativ cu pretenții teoretice se pot folosi termeni mai mult sau mai puțin definiți, prin implicarea conceptelor în cercetarea empirică acestea trebuie să dobîndească claritatea necesară identificării obiectelor reale ale căror proprietăți se sub-



sunează conținutului noțiunilor respective.

### 2.3. Statutul științific al sociologiei

Așa cu, am mai remarcat, nivelul teoretic al sociologiei nu strălucește prin virtuți demne de invidiat; de aici și o modestă poziție științifică a sociologiei. O privire superficială, aruncată fie din interiorul și mai cu seamă din exteriorul acestei discipline va surprinde o stare ce, în termeni foarte succinți și poate prea brutali, ar putea fi descrisă în felul următor. La nivelul teoriei generale, se constată o luptă ascutită între diferitele orientări, în ciuda faptului că actualmente în occident se manifestă o clară penurie de asemenea teorii care ar putea concura marxismul. În sociologia marxistă, dezvoltările teoretice sînt relativ timide, ele reducîndu-se în bună măsură la asimilarea unor concepte și scheme explicative utilizate cu mai mult sau mai puțin succes în alte cîmpuri teoretice. La nivelul teoriilor de rang mediu sau al schemelor explicative observăm o diversitate deconcertantă iar la cel al cercetării empirice s-ar putea vorbi chiar, dacă termenul nu e prea tare, de o adevărată junglă de date, de scheme explicative ad-hoc și de instrumente de investigație.

O analiză mai profundă poate pune totuși în evidență anumite principii și părți ce fac posibilă o unificare între diferite teorii, iar, la nivelul investigației empirice, rezultate similare, parțial verificabile. Vorbind despre aceste aspecte la Congresul de sociologie de la Evian, în 1966, reputatul sociolog american P. Sorokin (1967) asemuia sociologia cu un Ianus cu două fețe, din care prima (lipsa de unitate) este mai ușor de remarcat, din următoarele motive:

- se pleacă de la premisa că în alte științe nu ar fi diversitate și chiar discordanță a teoriilor;

- în sociologie apare cu pregnanță problema limbajului foarte diferit, care însă la o cercetare mai atentă se dezvăluie a fi mult mai uniform decît în aparență (mulți termeni se suprapun, total sau parțial);

- sociologii desconsideră adesea "principiul limitei", adică au tendința de a-și extinde propozițiile dincolo de domeniul lor de validitate;



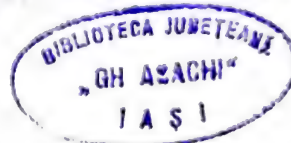
- sociologii neglijează aspectul multidimensional al realității sociale și exagerează o dimensiune.

În ciuda faptului că în ultimele decenii s-au făcut progrese însemnate spre o mai mare rigurozitate, atât în construcțiile teoretice, cât și la nivelul cercetării empirice, gradul de consistență al sociologiei se cere încă substanțial consolidat. Rămân, de pildă, valabile cuvintele spuse de un alt mare sociolog, E. Shils (1967), la același congres de la Evian; după a sa părere, una din deficiențele majore ale sociologiei rezidă în faptul că teoria nu a ajuns încă suficient de concretă pentru a veni în contact cu datele reale și, de aceea, nu e în măsură să-și încorporeze, la rândul ei, rezultatele analizei acestor date.

Care ar fi, în fond, explicația faptului că sociologia se găsește într-o asemenea stare? Încercările explicative n-au lipsit, ba dimpotrivă sînt atât de numeroase încît numai o analiză a lor ar umple probabil un tom. Fără pretenția de a fi sistematizat corect și, cu atât mai puțin, de a fi inventariat complet toate ideile emise în legătură cu această problemă, voi enumera unele din cauzele citate în literatura filosofico-metodologică pentru a explica statutul științific mai modest al sociologiei:

- a) complexitatea fenomenelor sociale, ceea ce înseamnă că orice analiză de efectuat și orice legitate de stabilit trebuie să implice un număr însemnat de variabile;
- b) legat de aceasta, dificultatea izolării unor aspecte autonome;
- c) efemeritatea multor fapte sociale, schimbările fiind evident mai rapide în societate decît în majoritatea celorlalte zone ale existenței;
- d) lipsa unui "contur" precis al anumitor fapte sociale, mai ales cele ce țin de conștiința socială;
- e) existența indivizilor umani în spatele oricăror fapte sau acțiuni sociale, ceea ce impune includerea în explicații, dacă nu a întregului univers psihic al acestora, cel puțin a unor elemente de bază, precum finalitatea acțiunilor individuale;
- f) observatorul se găsește "acufundat" în realitatea socială, fiind copleșit de aceasta (paradoxul ochiului care se auto-contemplă);

509.872





g) dificultăți legate de măsurarea multora (dacă nu a majorității) dintre aspectele vieții sociale;

h) dificultăți legate de posibilitățile efective de observare și, mai ales, de experimentare;

i) lipsa unor instrumente neutre, ca în fizică, de exemplu, ceea ce face ca în sociologie observatorul să fie, într-un fel, el însuși și instrument, de unde rezultă atât imperfecțiunea observației, cât și un adaos de subiectivitate;

j) condiționarea socială și culturală a cercetătorului;

k) legătura imediată a rezultatelor cercetării sociologice cu viața de zi cu zi a oamenilor, de unde posibile presiuni extra-științifice și manipulări din partea factorilor ce dispun de o asemenea putere;

l) lipsa unor informații primare consistente, obiective și acoperind intervale temporale lungi;

m) existența unei "sociologii vulgare";

n) informația fiind luată adesea de la indivizi umani nu prin observație externă, neutră, ci prin interogare, interviu, sociologul se ciocnește de un ecran interpus între el și realitate, fie în sensul că poate lua drept adevărate informații false (este mințit, pur și simplu), fie că trebuie să realizeze o interpretare mediată a realității. Este aici problema surprinsă foarte plastic de Borudieu și colaboratorii (1971), când afirmă că "este un blestem ca științele umane să aibă de a face cu un obiect care vorbește..." (p.192).

Cum se vede, multe dintre motivele invocate sînt legate unele de altele și s-ar cere, desigur, o mai amplă explicitare a lor și o analiză ceva mai aprofundată, dar spațiul nu-mi permite acest lucru.

Ceea ce doresc să mai adaug este faptul că, pe lângă factorii amintiți, factori intrinseci care țin de natura însăși a sociologiei, există și alții, exteriori, legați de funcția și rolul unei științe în societate, avînd o influență de netăgăduit asupra acesteia. Se știe că (și lucrul este mai evident în societățile moderne) activitatea științifică a căpătat pe măsura trecerii timpului un caracter de interes social tot mai important. Astăzi este din ce în ce mai greu de a întîlni cercetători ai

st  
pr  
ch  
ma  
tă  
ta  
pc  
bi  
cu  
le  
ve

ri  
tă  
st  
re  
pr  
pr  
ce  
ni  
te  
re  
de

ab  
pr  
ma  
pos  
mai  
de  
le  
și  
ven

soc  
tel



științelor empirice care să lucreze singuri, izolați și pe cont propriu. Munca în echipe, un anumit nivel minim de dotare tehnică, chiar și pentru științele sociale, pun în evidență dependența tot mai pronunțată a cercetării de factorii de decizie, de disponibilitățile bugetare pe care o societate sau alta le are pentru dezvoltarea științei. Dacă lucrurile sau așa, este limpede că pentru gospodărirea acestor mijloace - forțamente limitate - se impune stabilirea anumitor priorități în alocarea lor. Or, trebuie s-o recunoaștem, științele sociale sînt net handicapate în raport cu celelalte în accesul la asemenea mijloace, cel puțin din trei motive fundamentale:

a) Într-o anumită fază de dezvoltare a societății, prioritară apare dobîndirea unor cunoștințe și mijloace care să permită o rapidă dezvoltare economică. La început deci, cercetarea științifică este orientată spre acele domenii care permit creșterea productivității muncii, dezvoltarea bazei de materii prime etc. prin mijloace pur tehnice. Abia realizată o anumită fază în acest proces, se vor face simțite ameliorările pe care le pot aduce cercetările socio-umane. Așa de pildă, mi se pare cu totul evident că nici o organizare științifică a producției într-un atelier nu poate ridica productivitatea muncii în aceeași proporție ca introducerea unei mașini ce efectuează într-un minut munca unui om pe timp de o zi, să zicem.

b) Dat fiind statutul lor științific și natura fenomenelor abordate, științele sociale nu pot oferi soluții nici atît de prompte și nici atît de sigure ca cele tehnice. Introducerea unei mașini noi într-o întreprindere - ca să reiau același exemplu - poate fi ușor acompaniată de o serie de calcule de eficiență, tocmai pentru că asemenea calcule privesc un domeniu foarte strîmt de aspecte (economice, de cele mai multe ori). Disciplinele sociale - sociologia mai ales - se preocupă de fenomene foarte complexe și, ca atare, gradul de calculabilitate al consecințelor unei intervenții sociale este extrem de redus.

c) Există o serie de reacții negative față de științele sociale, mult mai evidente și chiar mai virulente decît față de altele și mult mai diversificate din perspectiva statutului social al



celor care le manifestă. Este clar, mai întâi, că cercetarea socială nu este totdeauna comodă pentru anumite persoane cu funcții de conducere. Un sociolog care "scormonește" în relațiile interpersonale sau în alte probleme ale unei întreprinderi trebuie să beneficieze de concursul unor cadre de conducere extrem de bine intenționate și cu o conștiință deosebită. Teama multora dintre astfel de oameni de a nu-și pierde funcția, dacă cercetarea sociologică dă la iveală anumite aspecte negative, este un lucru real de care trebuie să se țină întotdeauna seama.

Dar nu numai din atari motive pragmatice pot apare reacții negative la adresa cercetărilor sociologice. Există o credință destul de răspândită în rândul oamenilor că explicarea fenomenelor sociale, prin faptul că sînt mai ușor perceptibile, este o treabă la îndemîna oricui. Parafrazînd o cunoscută expresie, aș putea spune că toți indivizii cred că sînt născuți sociologi; fiecare are o explicație, mai mult sau mai puțin personală, pentru orice fapt, de la întîrzierea unui autobuz așteptat în stație și pînă la izbucnirea unui conflict armat între două state; fiecare face, este dispus - și într-un fel constrîns de însăși viața lui în societate - să facă previziuni asupra consecințelor actelor sale sau ale altora, fiecare este gata să ofere soluții pentru îndreptarea tuturor neajunsurilor.

Si nu este numai atît. O bună categorie de intelectuali, mai cu seamă din rîndurile literaților, consideră că rezultatele la care ajung științele sociale sînt minore în comparație cu complexitatea fenomenelor studiate și că literatura sau arta sînt capabile să dezvăluie lucruri mult mai profunde despre om și societate. Nu este locul să analizeze aici raportul dintre cele două domenii de activitate intelectuală - științele sociale și arta (mai ales literatura) - nici nu doresc să neg rolul pe care îl poate avea în dezvoltarea unor imagini ale realității sociale și nici nu țin să analizez cauzele disputelor numeroase dintre reprezentanții celor două sfere de activitate. Aș afirma doar că reproșurile făcute disciplinelor sociale sînt în mare măsură nedrepte, după cum inoportună este contrapunerea celor două genuri de activitate, care nu sînt deloc concurente, ci complementare și nici o cultură modernă nu-și mai poate permite să se dispenseze de una din ele.



În fine, mai există o reacție negativă vizavi de științele sociale, pe temeiul faptului că acestea, prin rezultatele consistente la care ajung - sau ar ajunge - pot servi pentru manipulare ori condiționarea oamenilor. Reproșul este copilăresc; este ca și cum ai vrea să interzici fizica deoarece dezvoltarea ei a condus la fabricarea armelor nucleare. Știința în sine nu este nici bună nici rea, numai utilizarea ei poate fi favorabilă sau defavorabilă omeniilor.

### 3. Domeniul metodologic

Întreg cursul de față va încerca în continuare să descrie unele aspecte ale acestui domeniu. În prealabil, se cuvin făcute două lucruri: definirea unor noțiuni sau, mai degrabă, precizarea sensului în care ele vor fi folosite aici și enunțarea câtorva idei despre valoarea reflexiei metodologice.

#### 3.1. Termeni

Diversitatea terminologică în domeniul de care ne ocupăm e de-a dreptul deconcertantă. Termeni ca metode, tehnici, proceduri, procedee, instrumente etc. se folosesc în accepțiuni foarte diferite, încât este greu să mai găsești ceva comun în acestea. Mai ales termenul de metodă este utilizat până la deplina uzare și pierderea oricărui conținut precis. Se vorbește de metoda dialectică, de metoda experimentală, de metode matematice sau statistice, de metode de eșantionare ș.a.m.d. Este greu să se pună ordine în acest haos printr-o convenție unanim acceptată și este dificil și pentru același om să fie absolut consecvent în folosirea termenilor respectivi. Nu-mi propun deci o clarificare a limbajului, ci voi face doar câteva precizări terminologice pe care voi încerca să le respect, pe cât posibil, până la sfârșitul cursului.

Prin domeniu metodologic înțeleg totalitatea mijloacelor, modalităților prin care o știință parvine la rezultatele sale. Aceste mijloace pot fi folosite fie în domeniul cercetării teoretice, fie în cel empiric și este clar că cele două moduri de activitate științifică uzează de mijloace diferite. În cursul de față,



voi prezenta unele din problemele tehnice ale cercetării empirice, care ajută la obținerea informației sociologice și la o prelucrare sumară a acestora.

Termenul de metodologie are o gamă largă de utilizări. În ceea ce mă privește, îi voi atribui un singur sens, și anume cel care derivă din etimologia cuvântului: teorie (discurs) asupra metodei. Mai exact, metodologia este o ramură științifico-filosofică ce are ca obiect de studiu domeniul metodologic al științei, urmărind stabilirea unor legități referitoare la valoarea de cunoaștere a diferitelor metode de cercetare. Vom avea deci o metodologie generală, corespunzând domeniului metodologic al tuturor științelor și metodologii particulare, pentru fiecare știință în parte sau pentru grupuri de științe apropiate. Vom putea vorbi astfel de o metodologie sociologică sau a științelor socio-umane.

Sensul acesta este destul de apropiat de cel propus de Lazarsfeld (1970), conform căruia metodologia nu este înțeleasă ca o disciplină normativă - cum se întâmplă de multe ori - chemată, să prezinte reguli de folosire a diferitelor mijloace de cercetare, ci una pozitivă, avînd drept sarcină de bază urmărirea sociologului în activitatea sa efectivă. Destul de apropiat, dar nu identic, deoarece o serie de funcțiuni pe care Lazarsfeld le atribuie metodologiei aparțin de fapt de domeniul teoretic al fiecărei științe. Dar iată care ar fi, după părerea sociologului american, temele majore ale preocupărilor metodologilor:

- a) delimitarea obiectelor, în sensul clarificării aspectelor decupate de sociologi din problematica vastă studiată;
- b) clarificarea sensului termenilor folosiți;
- c) explicarea tehnicilor de cercetare;
- d) punerea în relație a diferitelor tehnici de cercetare;
- e) sistematizarea rezultatelor empirice;
- f) formalizarea raționamentului (pp.258-266)

Al treilea termen fundamental ce se cuvine precizat este cel de metodă. Voi reține două dintre accepțiunile sale cele mai răspândite. Mai întîi, prin metodă se va înțelege o modalitate ge-



nerală de abordare a realității, un principiu fundamental ce ghidează întreaga desfășurare a unei cercetări. Astfel, se poate vorbi de o metodă structuralistă, funcționalistă, dialectivă, istorică, genetică, monografică, interdisciplinară etc. Se vede din înșiruirea de mai sus cât de eterogene sînt categoriile folosite, în sensul că ele nu sînt nici exclusive, nici de aceeași natură și nici de aceeași anvergură; unele din aceste tipuri de abordări derivă din concepția filosofică, altele din teoria sociologică iar altele sînt determinate de obiectul studiat.

În al doilea rînd, metoda poate primi un înțeles mai îngust, de modalitate efectivă de investigare a realității. Pornesc de la constatarea că științele naturii uzează de două metode de investigație empirică : observația și experimentul. Acestea, rămîn, la urma urmei, metodele de bază și ale sociologiei. Dar, avînd în vedere un specific cu implicații metodologice deosebite, și anume faptul că în științele socio-umane subiectul cercetător poate intra în relații de comunicare cu obiectul investigat, prin intermediul limbajului (comunicare ce poate lua diferite forme, de la o discuție față în față pînă la receptarea conținutului unor documente), observația ia unele forme ce se îndepărtează mult de ceea ce este în științele naturii, forme care, prin autonomizare, pot primi titlul de metode (specifice) ale științelor sociale.

La rîndul lor, metodele se pot subdiviza în funcție de tehnicile folosite, adică de operațiile concrete de înregistrare și manipulare a faptelor. În continuare, tehnica presupune, de cele mai multe ori, un instrument pentru a se putea realiza, cum ar fi, în cazul științelor sociale, chestionarul, ghidul de interviu, teste etc.

O clasificare a metodelor sociologiei, cu anumite subdiviziuni determinate de tehnicile folosite, ar putea fi următoarea:

a) Observația propriu-zisă

- exterioară (cînd cercetătorul nu se implică în viața comunității studiate);

- prin coparticipare (cînd se implică)

b) Interviul (dialogul față în față, urmărindu-se aprofundarea unor aspecte printr-o discuție relativ liberă)



c) Ancheta (caracteristica metodei este utilizarea chestionarului ca instrument de cercetare)

- directă (cînd chestionarul este completat de așa-numiții "operatori de anchetă" sau de persoane din echipa de cercetare)

- indirectă (autocompletarea chestionarelor)

d) Analiza documentelor

e) Experimentul

Termenul de procedee va fi folosit, de regulă, pentru a desemna operații și modalități de prelucrare a datelor unei cercetări empirice. Dar o utilizare riguroasă a lui în acest sens ridică numeroase probleme, dată fiind varietatea extrem de mare a mijloacelor de prelucrare a datelor, oferite mai ales de statistică. Astfel, termenul de procedeu ar trebui folosit pentru operații de anvergură foarte diferită, mergînd de la un simplu calcul de procent pînă la modele matematice sofisticate. De aceea - și pentru că există o destul de netă diferență între înregistrarea datelor pe teren și prelucrarea lor ulterioară - consider că nu se ridică nici o problemă dacă se reiau termenii de "metodă" și "tehnică", mai ales cu funcții stilistice, pentru variația limbajului, și în acest din urmă domeniu, atunci cînd nu există riscul nici unei confuzii asupra semnificatului.

În fine, metodica cercetării este concepută - într-un sens - a fi strategia concretă întrebuintată în cursul unei cercetări efective, incluzînd metodele, tehnicile instrumentele utilizate precum și alte aspecte de ordin organizatoric. În același timp, metodica cercetării este o disciplină cu caracter normativ care, bazată pe experiența de cercetare empirică realizată pînă în prezent, încearcă să degaje anumite reguli, norme etc. a căror respectare este necesară pentru asigurarea unei bune desfășurări a cercetării și pentru realizarea unei eficiențe cognitive sporite. Ea prescrie, de exemplu, regulile pentru întocmirea unui chestionar, precizează condițiile concrete în care anumite metode sau tehnici cu șanse mai mari de a da rezultate pozitive, evidențiază capcanele metodice în care cercetătorii cu mai puțină experiență pot cădea etc. Dată fiind finalitatea aceasta forată concretă a metodicii, rezultă că și aserțiunile sale au ca obiect niște aspecte foarte



concrete. Vreau să spun prin asta că metodică se preocupă nu atât de valențele cognitive ale metodelor - nici în sensul general al termenului de metodă, nici în cel mai restrâns din care a rezultat enumerarea de mai sus -, ci în primul rând de utilizarea efectivă a tehnicilor și instrumentelor de cercetare sau, mai exact, spus, se preocupă de metode doar prin prisma modului în care acestea se concretizează în tehnici și instrumente.

### 3.2. Valoarea reflexiei metodologice în sociologie

Este greu de găsit o altă știință despre al cărui domeniu metodologic să se fi scris mai mult decât despre cel al sociologiei. Situația aceasta a constituit nu o dată ținta unor critici serioase sau a unor ironii, prin care sociologii erau îndemnați să producă mai multe rezultate pozitive despre realitatea socială și nu numai discursuri nesfârșite despre valoarea cutărierii metode sau tehnici de cercetare.

Abordarea domeniului metodologic al sociologiei s-a făcut și se face, în general, din două unghiuri principale:

a) din cel al metodologiei și

b) din cel al metodicii.

În primul caz, se analizează valoarea de cunoaștere a diferitelor metode și tehnici; se discută probleme precum posibilitatea măsurării fenomenelor sociale, a explicației cauzale, a experimentării etc. Din păcate, se amestecă aici probleme de natură foarte diferită, genul acesta de discurs fiind alimentat, în egală măsură, de considerații filosofice speculative și de constatări pertinente desprinse din analiza riguroasă a cercetărilor efective, cum preconiza Iarasfeld. În al doilea, se prezintă, cum spuneam, regulile dobândite în practica cercetării concrete.

Cele două tipuri de abordare apar deseori îngemănate în una și aceeași lucrare metodologică și de aceea ele, în practică sînt destul de greu de separat. Insuși cursul de față, deși mă voi strădui, mai ales în secțiunea ce urmează, să mă axez cu preponderanță pe problemele metodice ale cercetării, este un complex de chestiuni metodologice (de exemplu, capitolul despre experiment), metodice (cel despre anchetă) și de probleme care aparțin ambelor



genuri dar și unei discipline autonome cum este statistica (ultimele părți ale cursului).

Proliferarea impresionantă a literaturii metodologice nu poate fi explicată numai prin cauze subiective, ținând de comoditatea cercetătorilor, care ar evita problemele dificile ale cercetării concrete și s-ar refugia în cîmpul speculațiilor metodologice, ci se datorează și unor factori mai generali, rezultînd din specificul disciplinei. În acest sens, se pot aduce cel puțin două argumente esențiale:

a) Mulți sociologi au considerat și consideră că reflexia metodologică reprezintă pentru disciplina pe care o slujesc o șansă de evitare a unei serii de ocolisuri sau impasuri ce se ivesc în mod firesc pe drumul dezvoltării normale a unei științe. Avînd în față știința cu un statut bine consolidat, sociologia poate folosi experiența acestora atît în domeniul construcției teoretice, cît și în cel al constituirii și utilizării arsenalului metodologic.

b) Specificul sociologiei și al celorlalte științe sociale, discipline care au de a face cu elemente precum scopuri, valori, acțiuni etc. ce nu se întîlnesc în științele naturii, a implicat, cum era și firesc să se întîmple, discuții numeroase, care au făcut să crească volumul scrierilor metodologice cu privire la mijloacele cele mai eficace de cercetare. S-a urmărit, în mod cu totul deosebit, clarificarea următoarelor două aspecte:

- măsura în care metodele tradiționale din științele naturii sînt utilizabile (sau pot fi făcute utilizabile) în cunoașterea fenomenelor sociale;

- raportul dintre mijloacele de cunoaștere specifice științelor sociale și cele din științele naturii.

Sînt probleme care au făcut să curgă multă cerneală și care sînt și astăzi, după mai bine de un secol, dezbătute cu aceeași pasiune de către filosofi și sociologi. Cele două direcții fundamentale, care s-au constituit și care au divizat lumea sociologică, poate mai evident - cel puțin la suprafață - decît înseși concepțiile teoretice, sînt, de regulă, numite monismul sociologic și dualismul sociologic. Prima direcție pleacă de la ideea unității metodelor de cercetare empirică indiferent de domeniul fenomenelor abordate, pe cînd a doua exacerbează distincția dintre științele



naturii și cele socio-umane (sau, cu alți termeni, ale culturii, ale spiritului etc.) atît pe linia obiectului, cît și a metodelor. Este clar că aceste concepții despre metodă interferează cu cele filosofice ale autorilor, cu construcțiile teoretice asupra fenomenelor sociale, de unde rezultă o mare diversitate în cadrul fiecărei orientări fundamentale.

Scopul cursului de față fiind unul tehnico-metodic, nu vom ataca interesantul cîmp de probleme de acest gen; am ieși astfel chiar și din sfera metodologiei, trecînd în cea a epistemologiei și filosofiei științei. (Pentru detalii, a se vedea, dintre lucrările mai recente, cea a lui L.Vlăsceanu, 1982). Subliniez doar, ceea ce s-a putut deja înțelege din cele spuse pînă acum, că opțiunea mea este pentru poziția monistă, prin urmare, pentru unitatea metodelor de cercetare empirică, cu singura precizare că acestea trebuie adecvate, în forma lor concretă de utilizare, la obiectele, fenomenele studiate și la raportul dintre obiect și subiect ce se instituie în procesul cunoașterii.

În măsura în care științele sociale aspiră la statutul de știință, ele trebuie să se supună rigorilor acestui statut. Una dintre cerințele elementare ale oricărei aserțiuni științifice despre un dat obiectiv constă în posibilitatea verificării acesteia de către orice alt cercetător. Or, o atare caracteristică devine posibilă numai în măsura în care se face apel la metode clare, standardizate de cunoaștere, așa cum utilizează de secole științele naturii. Esențialul, cred eu, în deosebirea dintre monismul și dualismul metodologic, nu constă în faptul că ultimul ar propune alte metode de cercetare în domeniul fenomenelor sociale, ci în acela că, odată cu repudierea metodelor naturaliste, este atacat, în esență, însuși spiritul științific al disciplinelor socio-umane.



## Cap. II. ETAPELE CERCETĂRII SOCIOLOGICE

Aparent, există o mare diversitate de opinii în literatura metodologico-tehnică privind etapizarea unei cercetări sociologice empirice, numărul fazelor ce trebuie realizate pe parcursul întregii investigații variind, în opinia autorilor, de la 3 la 15 sau chiar mai multe. Spun că această diversitate este doar aparentă pentru că, de fapt, aproape de fiecare dată sînt precizate destul de exhaustiv momentele principale ale cercetării, numai că ele apar cu statute diferite: de etape propriu-zise sau subetape. În fond, o discuție prelungită asupra acestei etapizări n-ar avea decît un caracter scolastic; important este, cum spuneam, ca momentele să fie enumerate și analizate, iar structurarea nu are, la urma urmei, alt scop decît să faciliteze înțelegerea discursului.

Orice autor de manual va trebui să opteze pentru o anumită formulă, desigur nu cu totul întîmplător, ci în virtutea unei concepții și perspective proprii - nu neapărat originale - asupra modului de realizare a cercetării. În ceea ce mă privește, cred că o formulă cu un caracter de mai mare generalitate - care să surprindă filmul desfășurării oricărei investigații sociologice și nu numai a anchetelor - ar fi cea care pornește inițial de la trei mari etape: pregătirea cercetării, culegerea informației și valorificarea informației. Acestea pot fi, la rîndul lor, subetapizate în funcție de metodicele concrete folosite în cercetările sociologice, dar, chiar și în cadrul lor, pot fi găsite elemente comune oricărei investigații.

Ultimele două etape sînt foarte strîns legate de metodele și tehnicile efectiv utilizate în practica cercetării și, prin urmare, amintitele elemente comune s-ar putea decela doar cu mare greutate; în schimb, etapa inițială pare a se preta mai ușor la o analiză sistematică din punctul de vedere al structurării operațiilor ce le incumbă. Din acest motiv, structura capitolului de față va fi deosebit de dezechilibrată, majoritatea comentariilor







referindu-se la faza pregătitoare a cercetării, mulțumindu-se ca la sfârșit să fac unele precizări legate de valorificarea informației, omițând complet, în acest moment, considerațiile despre cum legerea informației, care vor apare în capitolele dedicate metodelor de cercetare sociologică.

### 1. Pregătirea cercetării

Cred că sînt puține manuale și tratate de metode și tehnici care să nu insiste de o manieră cu totul deosebită asupra importanței pe care etapa pregătitoare o are asupra desfășurării efective a cercetării și asupra rezultatelor finale obținute. Chiar și fără asemenea insistențe, oricine poate să-și dea seama că rezultatul muncii sale depinde de gradul de perfecționare a instrumentelor utilizate, de cunoștințele pe care le are despre obiectul de studiat, de felul în care este pregătită și organizată din punct de vedere administrativ o investigație ce reunește adesea un număr mare de persoane etc. În ciuda tuturor avertismentelor și a evidenței, întîlnim, nu odată, și nu numai în cazul începătorilor în ale muncii sociologice, investigații pornite cu ochii închiși, cercetări pentru care se cheltuiește un volum imens de muncă și de mijloace materiale dar care ajung la rezultate minore tocmai din cauza neglijării pregătirii satisfăcătoare a lor. E vorba, de bună seamă, în cele mai multe cazuri, de anchete pe bază de chestionar, a căror simplitate aparentă poate duce foarte ușor în eroare. Carențele de construcție a chestionarului, de instruire a operatorilor, de alegere a eșantionului ș.a.m.d. vor avea implicații negative asupra eficienței anchetei, reducînd-o uneori aproape la zero.

Bineînțeles că pregătirea temeinică a unei cercetări constituie o condiție fundamentală cu caracter de principiu pentru orice studiu indiferent de disciplina științifică la care aparține. Dar, în același timp, pe lîngă remarcile de mai sus că există destule cercetări sociologice complet sau aproape complet ratate din lipsa unor preparative corespunzătoare, trebuie să înțelegem și faptul că nu toate investigațiile pot fi, obiectiv vorbind, la fel de bine pregătite. Deci gradul în care se realizează exigențele fazei la care mă refer depinde de foarte mulți factori și nu numai



de sociolog, adică de priceperea sau nepriceperea lui, de precau-  
țiunile pe care le ia sau nu, ori - de ce nu ? - de onestitatea  
sau lipsa lui de onestitate profesională. Vreau să spun că exis-  
tă cercetări în care o serie de condiții independente de sociolog  
nu permit realizarea tuturor momentelor pregătitoare: lipsa unor  
studii similare, deci nouitatea în abordarea fenomenului, care ne  
privescă nu numai de posibilitatea comparării datelor obținute dar  
și de utilizarea unor concepte deja folosite în cercetarea empirică,  
de instrumente standardizate, de ipoteze cu o verosimilitate ridi-  
culă etc. poate constitui o cauză ce conduce la rezultate mai pu-  
țin consistente. Astfel de investigații sunt numite uneori studii  
cu caracter explorativ spre a le deosebi de cele denumite cu ca-  
cter explicativ. Probabil că eticheta folosită pentru acestea  
din urmă nu este prea fericită aleasă, deoarece critic de modest ar  
fi un studiu, el nu poate renunța - fără a-și pierde statutul de  
cercetare științifică - la ideea explicării faptelor studiate.  
Oricum, distincția de mai sus pune în evidență faptul că în anumi-  
te situații caracterul explicativ este mai slab sau mai puternic,  
în funcție de posibilitatea realizării unei temeinice etape pregi-  
titoare a cercetării.

Pe de altă parte, sociologul se poate lovi de obstacole  
de timp, de bani, etc. - mai ales în cazul cercetărilor contractua-  
le - ceea ce-l conduce la realizarea unor investigații în pripă și  
fără mijloace materiale și umane corespunzătoare. Or, se poate  
afirma cu destul tenei că atunci când sociologul este presat de  
timp sau de lipsuri de mijloace materiale - dar mai ales de timp -  
el va renunța în primul rând și în cea mai mare măsură la o parte  
din activitățile reclamate de prima fază a cercetării, fiindcă  
lucrul acesta poate fi cel mai ușor de ascuns de ochii beneficiarilor  
lui.

Să trecem la prezentarea principalelor operații ce se  
cer realizate în etapa pregătitoare, acordând un spațiu mai impor-  
tant celor în care aspectele metodice sînt mai proeminente.

cărei  
esența  
rile  
fecțiun  
muncii  
tălor :

gares  
de lib  
finan  
de mar  
biecti  
țiale  
cînd c  
sau cu  
gitear

lucrar  
pentru  
bării  
deasă

de ni  
cît a  
mului

gică  
cerce  
condu  
lar d  
clar  
profi



### 1.1. Alegerea problemei (temei) de cercetare

Este normal că ne aflăm aici în punctul inițial al oricărei investigații. Din perspectiva cursului de față, chestiunea esențială ce trebuie subliniată rezidă în necesitatea circumscrierii corecte și cât mai exacte a problemei de cercetat. Orice imperfecțiune a acestei operații are consecințe negative atât asupra muncii de cercetare propriu-zisă, cât și, evident, asupra rezultatelor studiului.

Din alt punct de vedere, aș dori să menționez că deși alegerea temei de cercetare este în aparență un act ce depinde numai de liberul arbitru al cercetătorului, în realitate se dovedește a fi unul ce i se impune acestuia din urmă într-o măsură suficient de mare. Este normal ca în alegerea respectivă să intervină și subiectivitatea cercetătorului, preferințele, experiența și înclinațiile sale, dar, de obicei - pentru că sînt foarte rare cazurile cînd cercetătorii sînt cu totul independenți de anumite instituții sau organisme sociale - intervin cel puțin patru elemente constrîngătoare în alegerea temei de studiu:

(i) Tematica de cercetare a instituției în care sociologul lucrează trebuie să se orienteze spre probleme de interes major pentru societate. Această tematică este supusă, de regulă, aprobării forurilor superioare instituției respective, foruri care ghidează și orientează cercetarea în domeniul în cauză.

(ii) Cercetătorul dispune totdeauna de un anumit volum de mijloace materiale și este evident că el nu va putea aborda decât acele teme a căror investigare este posibilă în cadrul plafo-nului de mijloace pe care le are la dispoziție.

(iii) În al treilea rînd, aproape orice cercetare sociologică este o muncă de echipă. Este normal ca în funcție de numărul cercetătorilor, de profilul preocupărilor și specializării lor, conducătorul echipei va fi influențat să opteze pentru anumite teme. Iar din punctul de vedere al membrilor echipei, este de asemenea clar că fiecare dintre ei va fi într-o anumită măsură marcat de profilul colectivului din care face parte.



(iv) Un alt set de factori care pot influența alegerea problemei de cercetare este constituit din elemente precum: gradul de dezvoltare a diverselor ramuri ale sociologiei, tradițiile existente, apariția pe "piața sociologică" a unor teme la modă, bibliografia și datele statistice de care se dispune etc.

În cazul când sociologul lucrează în cadrul unei instituții neacademice, atunci cu siguranță că întreaga sa activitate de cercetare empirică va fi determinată de profilul instituției respective, deci domeniul de alegere se reduce simțitor. Sociologul de întreprindere va fi chemat să abordeze doar probleme care prezintă interes pentru acea întreprindere, preocupându-se de teme de sociologia muncii și sociologia industrială; cel încadrat la un centru de proiectare în domeniul sistematizării se va ocupa în mod precumpănitor de teme de sociologie urban-rurală, demografie etc.

### 1.2. Studiul bibliografiei

Probabil că pentru un fizician sau alt cercetător din științele mature o asemenea recomandare de a studia bibliografia înainte de a se apuca de o cercetare ar părea cel puțin superflua, dacă nu de-a dreptul ridicolă. Din păcate, în cadrul sociologiei ea trebuie făcută mereu și chiar subliniată din când în când și asta nu doar pentru faptul, clar și îndeobște cunoscut, că nu te poți angaja într-un studiu serios fără a cunoaște ceea ce au făcut și alții în domeniul respectiv. Situația sociologiei este de așa natură încât permite saltul peste această etapă (sau, mai corect spus, lasă iluzia că s-ar putea trece peste ea). Cunooscând precaritatea multor cercetări, sociologul poate exagera uneori în sensul de a considera că tot ceea ce au făcut alții este lipsit de valoare și deci se poate dispensa de experiența lor. O altă față a aceleiași chestiuni se poate manifesta chiar atunci când sociologii cunosc rezultatele confrăților lor, dar în cercetările proprii, dacă nu le ignoră, încearcă să le "depășească" printr-un exces de originalitate. Probabil că avem de a face aici cu o influență nefastă a pregătirii în general filosofice a sociologilor noștri, influență concretizată în iluzia că originalitatea se confundă cu valoarea. Din nefericire - sau fericire - în știință lucrurile



nu stau deloc așa, pentru că aici aspectele cu totul originale sînt mai degrabă excepții; regula o constituie munca asiduă de culegere și interpretare a informației în cadre bine constituite, muncă ce nu se poate realiza decît printr-o cunoaștere exactă și clară a ceea ce s-a realizat, printr-o continuitate a preocupărilor și metodelor de lucru.

Prin urmare, studiul bibliografic permite sociologului o ameliorare a rezultatelor muncii sale din perspectiva mai multor factori, dintre care amintesc:

- a) delimitarea precisă a obiectului cercetării;
- b) utilizarea unor teorii sau scheme explicative a căror validitate a fost testată;
- c) folosirea unor ipoteze de lucru rezultate dintr-o cunoaștere prealabilă a domeniului;
- d) aplicarea unor metode, tehnici, instrumente și procedee care au condus deja la rezultate consistente în domeniul respectiv.

Una dintre consecințele de cea mai mare valoare pentru dezvoltarea sociologiei pe care o are studierea aceluiași fenomen cu metode, concepte și cadre teoretice similare o constituie comparabilitatea rezultatelor studiilor sociologice empirice. Este realmente o problemă de o deosebită însemnătate, fără rezolvarea căreia este imposibil să se îndeplinească condiția de bază a dezvoltării unei științe; cumulativitatea cunoștințelor dobîndite. Din păcate, sociologii păcătuiesc, cum spuneam, de excesul de originalitate, au mania "delimitărilor" permanente de preocupările altora și asta nu numai în cîmpul cercetării teoretice dar și al celei empirice. De aici munții de date empirice culese de aproape trei sferturi de secol de sociologie de teren regulată, din care însă nu au apărut multe concluzii de mare amplitudine.

### 1.3. Precizarea conceptelor

Este o operație strîns legată de precedenta, căci ea nu se poate realiza fără cunoașterea încercărilor - reușite sau nereușite - ale predecesorilor. Despre importanța conceptelor, în general într-o cercetare științifică, s-a vorbit atît de mult încît



nu are rost să mai insist aici asupra ei. Menționez doar că, de regulă, conceptului i se atribuie trei funcții importante pentru cercetarea empirică:

- să organizeze percepția
- să ghideze observația și
- să prezică fapte noi.

Așa cum am mai spus-o, în sociologie avem de a face o penurie de concepte bine precizate. Există o mare ambiguitate a termenilor, împrumutați în imensa lor majoritate din limbajul comun. Din această cauză, utilizarea lor fără o definiție prealabilă devine, dacă nu imposibilă, în orice caz ineficientă în munca de cercetare. Desigur că nici definirea de fiecare dată în mod original a conceptelor nu este o modalitate prea elegantă și economică pentru dezvoltarea unei științe.

Dacă s-ar putea măsura gradul de dezvoltare al științelor nu cred că ar fi greu să se pună în evidență o corelație negativă puternică între acest indicator și numărul conceptelor folosite de diferite științe. Iar în sociologie, trebuie s-o recunoaștem, există o inflație de noțiuni. De pildă, ca să reiau un exemplu vehiculat în literatura metodologică, în 1931 un autor american, pe nume Eubank, a întocmit o listă cu noțiunile principale folosite în 8 lucrări de sociologie cu caracter general, ale unor autori diferiți, listă care cuprindea în total 146 de concepte. Dintre acestea nici unul nu a putut fi regăsit în toate cele 8 lucrări; în 7 lucrări s-a regăsit unul singur comun, în 6 lucrări - două iar în 5 lucrări - opt. Din totalul de 146 noțiuni, 83 erau specifice câte unei singure lucrări. Si nu sînt convins că situația s-a ameliorat radical de atunci și pînă azi. Chiar și în cazul cînd unele noțiuni au o circulație mai largă, sensul și semnificația lor variază sensibil de la autor la autor.

Se impune deci cu necesitate faptul ca la începutul oricărei cercetări sociologice de teren să se definească principalele concepte cu care se lucrează, iar dacă este posibil această definiție să se facă pe linia uniformizării limbajului, deci în mod similar altor cercetări, pentru comparativitatea datelor. Este necesară definirea precisă la început deoarece de semnificația acordată conceptelor depinde în mod hotărîtor întregul proces de culegere, sistematizare, prelucrare și analizare a informației.



#### 1.4. Operaționalizarea conceptelor

Pentru orice știință empirică, una dintre problemele sale fundamentale este aceea de a stabili modalitățile prin care se realizează legătura dintre universul teoretic și realitatea concretă, să-i zicem, reflectată. În cadrul acestei chestiuni generale, un loc central îl ocupă problema operaționalizării conceptelor, ca ansamblu de operații prin care însușirile definitorii ale noțiunilor pot fi identificate sau măsurate în universul empiric. O abordare teoretică de profunzime a acestei probleme ne-ar duce foarte departe la aspecte de epistemologie, filosofia și logica științei, metodologie generală etc. În scopul înțelegerii activității practice de operaționalizare, voi schița doar câteva repere.

Mai întâi - și iarăși fără a intra în chestiuni filosofice adânci, în speță de teoria definiției - se cuvine observat că în sociologie apar cu cea mai mare frecvență două tipuri de noțiuni: unele, pe care să le numim de tip atributiv și altele, de tip variabil. Primele sînt cele studiate pe larg de logica clasică bivalentă, prin punerea în evidență a sferei și conținutului noțiunii. Ele presupun, în fond, prezența unui univers de indivizi empirici care poate fi decupat în două submulțimi: cea formată din "obiectele" ce poartă însușirea (sau însușirile) cerută de conținutul noțiunii - și formînd sfera acesteia - și cea formată din cele care nu posedă trăsătura respectivă. Așa de exemplu, în cercetările noastre avem de a face cu noțiuni ca "muncitor" (membru al clasei muncitoare), "activ", "retribuit", "ateu", "căsătorit" ș.a.m.d. toate acestea fiind evident de tip atributiv.

Noțiunile de tip variabil presupun, de asemenea, un univers de "obiecte" empirice (care, ca și în cazul precedent, pot fi indivizi umani, grupuri sociale, instituții, comunități sau relații umane, acțiuni sociale etc.) dotate cu o anumită trăsătură ce nu se mai evidențiază doar prin cele două alternative - prezență / absență -, ci printr-o variație graduală în intensitate. Așa sînt conceptele de satisfacție în muncă, de integrare socială, de coeziune a unei colectivități și multe, foarte multe altele. Chiar dacă unele noțiuni de tip variabil pot fi transformate, prin



simplificare, în altele de tip atributiv (vorbind despre indivizi satisfăcuți de muncă sau nesatisfăcuți, despre integrați sau ne-integrați etc.), diferența este totuși esențială, mai ales în problema ce ne interesează aici.

Revenind acum tocmai la această problemă, vom mai observa că literatura metodologică face distincția dintre așa-numitele concepte ideale, abstracte, latente, neoperationale etc. "care nu au referințe observaționale (empirice) nemijlocite" și altele empiric referențiale, manifeste, concrete, operationale etc., ce "definesc entități direct observabile" (Vezi L.Vlăsceanu, 1982, pp.258-259). O astfel de delimitare, deși rigidă, ca orice dihotomie, are meritul de a sugera foarte clar problematica operaționalizării conceptelor. Intr-adevăr, un concept de tip variabil ca "vârsta" poate fi imediat perceput în realitatea obiectivă (fie întrebând individul în cauză sau persoane apropiate de acesta, fie consultând certificatul de naștere sau alt document potrivit). Dimpotrivă, satisfacția muncii sau calitatea vieții unei persoane sînt practic nedecelabile printr-o operație empirică atît de simplă. Cum se procedează în asemenea cazuri ?

Dacă ne-am apleca asupra unor noțiuni obișnuite, utilizate în mod curent în limbajul comun, vom observa că, de fapt, fiecare dintre noi realizează operaționalizări fără să-și dea seama atunci cînd distinge, de exemplu, un bărbat de o femeie, o relație de rudenie de una de serviciu, un profesor de un elev, o stare de tristețe de una de veselie etc. De multe ori oamenii nu au cunoștințe prea clare despre conținutul unor asemenea noțiuni, ceea ce nu-i împiedică să le folosească totuși în mod curent și cu relativ succes (adică fără să intre prea des în situații dilematice). Este probabil foarte greu pentru un om obișnuit să definească în mod acceptabil, de pildă, noțiunea de tristețe, dar nimeni nu se îndoiește de capacitatea sa de a diagnostica tristețea (cel puțin în rîndul persoanelor ce-i sînt apropiate) prin semne care i se par absolut semnificative în comportamentul semenilor săi. Mai mult, chiar dacă e ușor de definit o noțiune, în practica zilnică nu este necesar - și nici comod - să se folosească în operațiile de identificare tocmai trăsăturile definitorii. O serie de alte aspecte, mai ușor de sesizat, sînt suficiente pentru a

se  
voc  
tot  
și

de  
re  
luc  
ca  
col  
di  
ap  
va  
ma  
ur  
da  
al  
an  
ur  
st  
un  
vi  
or  
de  
re  
te  
in  
in  
te  
di  
tr

ri  
pe  
23

de  
fi





se ajunge la același rezultat. Astfel, felul de a se îmbrăca, vocea, coafura, forma corpului etc. sînt elemente care în cvasi-totalitatea cazurilor permit să se facă distincția între o femeie și un bărbat, cel puțin de la o anumită vîrstă în sus.

Conceptele mai complexe, care urmăresc aspecte tot atît de reale ca și cele din exemplele amintite aici, se definesc, de regulă, prin trăsături mai generale, fiind rezultatul unor prelucrări și rafinări ale altor concepte mai simple. În astfel de cazuri, apropierea lor de realitate nu se poate face decît prin constituirea unor trăsături simple care să poată fi identificate direct în empirie. Este vorba de un set de însușiri, pentru că e aproape imposibil de găsit o trăsătură simplă care să fie echivalentă cu cea complexă din definiție; numai prin "însușirea" mai multor aspecte parțiale se poate spera să se ajungă la scopul urmărit. Mai mult, gradul înalt de complexitate a unor concepte se datorează și particularității lor de a acoperi dimensiuni distincte ale realității. Astfel, dacă vorbim de integrarea studenților din anul I în viața universitară, trebuie să avem în vedere cel puțin următoarele aspecte: integrarea în colectivul anului (grupele de studiu, adaptarea la stilul și cerințele activității profesionale universitare, integrarea în viața organizației ASC, adaptarea la viața de cămin (dacă e cazul), adaptarea la stilul de viață al orașului (cel puțin pentru cei nou veniți în oraș) etc. Este evident că aceste paliere apar efectiv ca dimensiuni distincte, deoarece ele sînt practic independente unul de altul (un student poate fi considerat perfect integrat în raport cu unul și totuși neintegrat în raport cu altul) și de asemenea, este normal ca în încercarea de evaluare sintetică să se țină seama de toate aceste dimensiuni. Pentru cercetarea concretă este nevoie ca fiecare dimensiune aleasă să fie, la rîndul său, operaționalizată, adică transpusă în indicii empirici.

Dar să-l urmărim mai bine pe P.Lazarsfeld, care, cu claritatea-i recunoscută, sintetizează etapele ce se cer parcurse pe drumul ce duce de la concept la studiul empiric (1971, pp.20-23), adică tocmai esența operaționalizării.

(1) Inițial avem o imagine sau o reprezentare destul de vagă a ceea ce înțelegem printr-o noțiune, imagine rezultată fie din decelarea unor trăsături comune subiacente mai multor



fenomene dispersate, fie din încercarea de a explica anumite constante.

(ii) A doua etapă constă în specificarea conceptului. Ea consistă în descompunerea reprezentării în componente, ce constituie diverse aspecte sau dimensiuni ale conceptului. Astfel de componente se deduc uneori din analiza teoretică a conceptului iar alteori rezultă din analiza empirică a corelațiilor observate.

(iii) A treia etapă se referă la alegerea indicatorilor. Indicatorul nu este altceva decât acea însușire simplă care poate fi detectată direct în realitatea empirică. De exemplu, dacă observăm că un om frecventează regulat biserica, acesta va fi un indicator destul de sigur al religiozității persoanei respective. Indicatorii pot fi, evident, de genuri foarte diferite, fie în funcție de modalitatea de conectare cu realitatea, fie după tipul de informație pe care ei o furnizează (de la specificări de genul da/nu până la valori numerice, rezultat al măsurării).

△ (iv) A patra etapă este cea a formării indicilor. Indicii empirici, cu ungrad mai mic sau mai mare de generalitate, urmăresc să surprindă sub o formă numerică concentrată informația conținută într-un set de indicatori. Dacă o dimensiune oarecare este operaționalizată printr-un singur indicator ce permite măsurarea, atunci se obține direct un indice pentru aspectul în cauză (de pildă, media valorilor individuale). Dacă sînt mai mulți indicatori, atunci sau se traduce fiecare într-un indice și aceștia din urmă se "condensează" apoi în unul singur (printr-o procedură de medie, de exemplu) sau se combină indicatorii între ei astfel încît să se ajungă la o singură valoare finală - indicele dimensiunii. În sfîrșit, ultima etapă de imaginat ar fi sintetizarea indicilor dimensiunilor în unul singur, general, capabil să reflecte întreaga problematică a conceptului. S-ar ajunge astfel la indici de calitate a vieții, de satisfacție, de integrare etc.

Desigur că tehnologiile de obținere a unor asemenea indici sintetici sînt destul de complicate, problemele metodologice și tehnice ridicate de operația respectivă nu sînt ușor de rezolvat. Există o literatură destul de întinsă legată de problematica indicatorilor și indicilor, dar în cursul de față nu putem zăbovi



mai mult asupra ei, mai ales acum în partea de debut. Astfel de chestiuni ar putea fi cu mai mult folos abordate după introducerea metodologică și statistică ce constituie, în fapt, obiectul întregului nostru curs. Ceea ce se poate face aici cu destul folos este să se realizeze o clarificare a lucrurilor prin exemple concrete.

Intr-o cercetare de sociologie industrială, C. Zamfir și colaboratorii (1980) au urmărit, printre altele, operaționalizarea noțiunii: "satisfacția muncii". Sistemul de indicatori folosit în acest scop a cuprins o baterie de 8 întrebări din chestionarul folosit, întrebări care cereau subiecților să exprime gradul de satisfacție încercată în legătură cu următoarele aspecte:

- profesiunea,
- întreprinderea,
- componența grupului de muncă,
- șeful direct,
- conținutul muncii,
- promovarea efectivă (pînă în prezent),
- posibilitățile de promovare (în viitor),
- retribuirea

Întrebările erau de genul: "În general, cît de satisfăcut sînteți de persoanele din grupul dv.?", cu răspunsuri mergînd pe o scală cu 7 trepte, de la foarte puțin la foarte mult, fiecare treaptă primind o valoare numerică de la 1 la 7. Din cumulara valorilor atribuite de fiecare individ la cele 8 întrebări, s-a obținut o cifră totală cu rolul de indice global de satisfacție.

După prelucrarea datelor și efectuarea unor analize statistice, s-a ajuns la concluzia că indicatorii folosiți tind să circumscrie trei dimensiuni ale conceptului:

- a) satisfacție în raport cu relațiile din colectivul de muncă;
- b) satisfacția cu conținutul și cadrul de desfășurare a muncii;
- c) satisfacția cu beneficiile personale asociate muncii.

Este evident că o asemenea procedură de construcție a dimensiunilor unui concept este total dependentă de natura și numărul indicatorilor aleși (în cazul de față, întrebările chestionarului). Rezultatul operației va fi variabil, de la cercetător la cercetător.



cetător, neexistînd garanția depistării "adevăratelor" dimensiuni ale conceptului. Dar nici procedeul invers - de stabilire a priori a dimensiunilor - nu este scutit de pericolul subiectivismului. În absența unei teorii clare, o analiză, oricît de atentă ar fi ea, nu poate garanța identificarea tuturor dimensiunilor unui concept. Eventuala consolare că o analiză ulterioară a indicatorilor aleși va confirma sau infirma dimensionarea apriorică (deci ne va ajuta în aprecierea validității sale) este pur și simplu o iluzie. Dacă dimensiunile sînt dinainte date, numai un cercetător lipsit de o minimă experiență (ca să nu zic de inteligență) va alege indicatori ale căror corelații să fie mai slabe între perechile formate din indicatorii ce intră într-o aceeași dimensiune, decît între perechile aparținînd la dimensiuni diferite. (În general, identificarea dimensiunilor a posteriori sau validarea celor introduse a priori se face cu ajutorul unor proceduri bazate pe coeficienții de corelație, coeficienți care trebuie să aibă valori mai ridicate între variabilele cuprinse în aceeași dimensiune decît între cele provenite din dimensiuni distincte).

Oricum s-ar proceda, este limpede că un concept odată operaționalizat, odată specificate și urmărite anumite dimensiuni ale sale, rezultatele cercetării empirice vor descrie numai aspectele reținute de dimensiunile și indicatorii utilizați. De aici rezultă cu claritate că extrapolarea constatărilor cercetării la întreaga semnificație a conceptului este ilicită și că, prin urmare, de cele mai multe ori, investigația empirică - tocmai prin surprinderea unui cîmp de dimensiuni - sărăcește, simplifică realitatea. Esențial aici este de a găsi o formulă simplificatoare acceptabilă, în sensul că vor fi totuși surprinse elemente esențiale și că ceea ce se pierde printr-o astfel de sărăcire este înzecit compensat de cîștigurile obținute prin precizia măsurărilor, profunzimea analizelor etc.

Așezarea problemei operaționalizării conceptelor în rîndul elementelor de bază în pregătirea cercetării și sublinierea importanței acestei probleme pentru eficiența investigației nu înseamnă acceptarea poziției filosofice cunoscută sub numele de operaționalism. Această concepție - o variantă a neopozitivismului - susține, în esență, ideea că definirea noțiunilor se reduce la specificarea operațiilor efectuate în universul datelor empirice. Caracterul său

rela-  
filo-  
mul  
meto-

int  
In  
lui  
sti  
găsi  
ții  
ipa  
re  
at  
in  
co-  
ze  
ru  
in  
te  
re  
de  
da  
cr  
te  
ta-  
la  
de  
ții  
De  
la  
ex-  
tun



relativist și subiectiv a fost pus cu claritate în evidență de filosofia marxistă, dar, ca și întreg pozitivismul, operaționalismul a avut și el o influență semnificativă asupra perfecționării metodelor de cunoaștere în științele socio-umane.

### 1.5. Stabilirea ipotezelor de lucru

Ne găsim și aici în fața unei probleme ce se află la interfața epistemologiei, filosofiei științei și metodologiei. În toate aceste domenii s-a abordat pe larg conținutul conceptului de ipoteză, precum și rolul și locul acesteia în cunoașterea științifică. Ceea ce doresc să menționez aici este o idee ce-și găsește loc adesea în lucrările metodologice. O investigație științifică nu se realizează niciodată în afara sau în lipsa unor ipoteze. Este o iluzie faptul că ipotezele ar lipsi din cercetare; ele există sub o formă inconstientă sau nespecificată chiar atunci când cercetătorul se străduiește să studieze o realitate în mod cât mai "obiectiv", fără "idei preconcepute". De exemplu, construcția oricărui chestionar se bazează pe o mulțime de ipoteze. Faptul că se pune o întrebare și nu alta înseamnă că cercetătorul a ales ca probabilă o legătură între fenomenele X și Y și nu între X și Z. Faptul, mai general, că se adoptă anumite metode și tehnici de cercetare se bazează tot pe ipoteze legate de specificul realității studiate ș.a.m.d. Deci, practic, orice operație, oricât de mărunță ar fi, este până la urmă impregnată, susținută și ghidată de anumite ipoteze.

Concluzia care se degajă este evidentă. Decît să se lucreze cu asemenea ipoteze neconstientizate, deci necontrolate, este preferabil ca înaintea oricărei cercetări să se facă un inventar cît mai complet al ipotezelor implicate. Ar mai fi de adăugat la aceasta că ipotezele nu apar numai înaintea muncii empirice, deci nu numai în legătură cu datele culese, ci și după etapa obținerii informației, atunci cînd se face interpretarea datelor. De pildă, dacă în etapa inițială presupunem existența unei corelații între două fenomene, în final, dacă respectiva corelație există, se vor emite ipoteze legate de raportul mai adînc de legătură între fenomene. Astfel de ipoteze pot fi testate empiric în



cadrul cercetării sau pot rămâne deschise și deci apar ca rezultat al unei investigații.

#### 1.6. Alegerea metodelor și tehnicilor

Reprezintă, desigur, un alt moment important al cercetării, care se realizează practic simultan cu alegerea problemei de cercetat, deoarece fiecare obiect antrenează tehnici de investigație specifice sau, mai bine spus, metodele și tehnicile se cer particularizate și concretizate în funcție de tema de cercetare. Am folosit pluralul în subtilul de mai sus deoarece adeseori în investigațiile sociologice se recurge la o îmbinare de metode și tehnici. Metoda anchetei, de exemplu, poate fi realizată în variante tehnice diferite și, în plus, ea poate fi și este adeseori însoțită de o analiză a documentelor, de observație sau interviu.

În afara temei de cercetare, arsenalul metodico-tehnic utilizat se mai alege desigur și în funcție de concepția teoretico-metodologică a cercetătorului, de experiența sa, adică de felul în care se simte capabil să mînuiască diferite, metode, tehnici și instrumente de cercetare. De asemenea, nu trebuie neglijată nici influența mediului sociologic general, care, în diferite momente istorice, impune recurgerea la metodici specifice de cercetare. În fine, nu se pot omite într-o asemenea enumerare nici mijloacele materiale și umane precum și timpul avut la dispoziție de cercetător. Aceste mijloace externe constrîngătoare își fac simțită influența, după cum se vede, în aproape toate elementele ce constituie esența desfășurării investigației sociologice. Oricît de mari ambiții ar avea un cercetător, dacă nu are la dispoziție bani sau personal calificat, nu va putea recurge la tehnologii scumpe sau care cer o pricepere deosebită în culegerea datelor. Probabil că acesta este și unul dintre motivele care au impus într-o vreme ancheta ca metodă privilegiată în studierea fenomenelor sociale.



### 1.7. Construcția instrumentelor

Este evident, un aspect ce decurge logic din cel anterior, instrumentul fiind acela care concretizează, materializează o tehnică oarecare. Prea multe lucruri generale de natură metodică nu se pot spune nici la acest punct, deoarece fiecare instrument de culegere a datelor ridică probleme specifice de întocmire. Despre instrumentul cel mai utilizat în sociologie - chestionarul - o să vorbim pe larg în capitolul destinat anchetei, întrucât o bună parte din problematica respectivei metode este legată de întocmirea, aplicarea și prelucrarea chestionarului.

Voi aminti aici doar cele două cerințe, fundamentale pe care trebuie să le îndeplinească, de fapt, orice instrument, indiferent de tipul de investigație empirică: fidelitatea și validitatea. Pe scurt, vom înțelege prin fidelitatea unui instrument capacitatea sa de a reproduce în mod constant anumite însușiri constante ale fenomenului. Validitatea este proprietatea instrumentului de a dezvălui cercetătorului ceea ce acesta urmărește să obțină în cercetare. Această din urmă cerință este mai "tare" decât prima, în sensul că un instrument poate să fie fidel fără a fi valid. De exemplu un test poate să dea rezultate identice aplicat la grupuri identice, dar să nu măsoare ceea ce-și propune.

În speță, referindu-ne la validitatea unui chestionar, se înțelege că, pentru realizarea ei, aspectul realității abordate trebuie surprins cât mai exact și mai complet posibil în seturile de întrebări, adică acestea trebuie să fie relevante pentru problema investigată, iar mijloacele de înregistrare - scalele, de exemplu - să surprindă de o manieră cât mai precisă informația culeasă. Această adecvare a instrumentului la obiectul științific și garantează o anumită formă de validitate, numită, de obicei, validitate externă. Se poate vorbi, în egală măsură, și de una internă, ce derivă din construcția instrumentului. Implicând, în principal, fidelitatea acestuia: informațiile obținute să poată fi univoc clasificate și, dacă e vorba de chestionare, teste etc., termenii folosiți să fie univoc înțeleși de către fiecare operator sau subiect.

Oricum, problema validității instrumentelor în cercetarea sociologică rămâne una dintre chestiunile de cea mai mare importanță,



deoarece instrumentul nu numai că reprezintă o sursă principală de erori, dar, de modul în care fenomenele sînt traduse în indicatori, de calitatea scalelor prin care se evaluează însușirile și de alte calități de construcție a instrumentului depinde calitatea rezultatelor, consistența acestora, în fond calitatea cunoștințelor dobîndite. Cercetările metodologice mai recente au pus în evidență unele modalități de testare a fidelității și validității instrumentelor sociologice. Din păcate, ele au încă o eficiență scăzută și, chiar și așa, putem spune că apelul făcut la ele de către sociologii noștri nu este totdeauna satisfăcător, pentru că nu rareori aceste proceduri sînt insuficient înțelese. Întîlnim destul de des formulări de genul: "consistența internă a chestionarului a fost demonstrată de..." sau "instrumentul posedă un înalt grad de validitate" etc., deci aprecieri foarte categorice apărute în urma unor operații formale ce reflectă puțin din fondul chestiunii, formulări care trebuie mereu privite cu rezervă și, aș zice, cu atît mai multă rezervă cu cît calitățile instrumentului sînt mai puternic apreciate.

#### 1.8. Delimitarea universului de studiu și a eșantionului

E vorba aici, mai întîi, de a decupa din realitatea obiectivă totalitatea obiectelor purtătoare de însușiri, obiecte ce constituie universul material al cercetării, aria, zona asupra căreia se vor extinde rezultatele investigației. Operația este legată foarte strîns de problema definirii conceptelor și de cea a construcției indicatorilor empirici.

Delimitarea riguroasă și eficientă pentru cercetare a populației sau universului de unități de studiat este o operație extrem de dificilă, ridicînd multe probleme de ordin practic și teoretic pentru investigatorul realităților concrete. Într-adevăr, acesta nu se poate mulțumi cu formulări generale sau abstracte, care pot oferi, în principiu, o imagine asupra cîmpului de investigat, dar care atunci cînd sînt puse efectiv la lucru se dovedesc incapabile să ofere criterii exacte de delimitare.

Să dăm un singur exemplu, aproape banal, dar care va oferi o imagine destul de concludentă asupra dificultăților cu care



se confruntă sociologul practician. I se comandă, să zicem, de către un studio de radio regional să testeze opiniile locuitorilor unui oraș despre programele postului de radio respectiv. Care va fi populația ce formează obiectul (material) al studiului? Desigur că după o prealabilă definire foarte clară a noțiunii de "locuitor al orașului X" va trebui să se elimine din populația astfel obținută toți locuitorii sub o anumită vîrstă (n-are sens să se pună problema pentru sugari). Pe de altă parte, opinia despre postul de radio în cauză nu poate fi cerută decît celor care, în general ascultă radioul și, în particular, celor care acordă o atenție relativ regulată emisiunilor postului nostru. Dar cum s-ar putea determina, de exemplu, într-un oraș de talia Clujului, ca să nu vorbesc de altele mai mari, cît de cît exact populația care ascultă programele de radio. Ne va duce gîndul, desigur, la faptul că, pentru a asculta, omul trebuie să fie ori în posesia unui aparat de radio, ori să aibă acces la el. Cît de greu este să se depisteze chiar și numai posesorii aparatelor de radio oricine își poate imagina; evidențele aflate la oficiile postale sînt destul de incomplete: există indivizi care nu declară posesiunea receptorului, așa după cum există alții în evidență - desigur mai puțini - cu aparate de radio stricate sau înstrăinate de o anumită vreme, fiind deci în imposibilitatea de a se pronunța asupra emisiunilor din ultima perioadă, chiar dacă ar fi fost ascultători ai postului respectiv. Problema accesului efectiv la radioreceptoare este și mai greu de rezolvat căci într-o familie poate exista un radio dar unii dintre membri să nu asculte, tot așa cum alte persoane care nu au în locuință receptoare să asculte în altă parte (la locul de muncă, la prieteni etc.). Practic, ar fi nevoie deci, numai pentru delimitarea populației de radioascultători, fără specificarea posturilor frecventate, să se facă o investigație prealabilă în orașul X. Nu este deci de mirare că de multe ori, în cazuri ca cel invocat mai sus, universul cercetării se conturează efectiv în cursul acesteia, lucrul posibil atunci cînd se pornește de la un univers mai larg (incluzîndu-l pe al nostru), univers ce oferă indicatori de delimitare foarte precizi.

Eșantionarea se folosește în situațiile în care populația este extinsă și nu este nici posibil, nici eficient să se studieze



în totalitatea sa. O subpopulație aleasă cu grijă, pe baza unor principii statistice derivate din teoria matematică a probabilităților sau pe temeiul unor cunoștințe dobândite în practică, poate suplini în bune condiții cercetarea universului întreg. Asupra acestei chestiuni, deosebit de importantă pentru investigația sociologică, voi reveni mai pe larg în ultima parte a cursului, arătând atât modalitățile practice de alegere a eșantionului cât și semnificația rezultatelor obținute în cercetările selective.

### 1.9. Organizarea acțiunii practice de cercetare

Intrucât cercetările sociologice se efectuează cel mai adesea cu un personal destul de numeros și pot avea întinderi temporale și extinderi teritoriale destul de însemnate (de unde și denumirea, folosită uneori, de "campanii" sociologice), este clar că pentru buna desfășurare a acțiunii trebuie să se depună un efort organizatoric deosebit. Numai cine a organizat efectiv o campanie sociologică cu cel puțin 20-30 de oameni în echipă poate aprecia foarte exact dificultățile ce se cer depășite în asemenea situații. Aceasta nu înseamnă că ele nu pot fi totuși sugerate, ceea ce voi încerca să fac, de o manieră foarte succintă, în cele ce urmează.

Mai întâi, se impune o riguroasă planificare în timp a desfășurării studiului, în care să apară foarte clar corelate toate etapele și operațiile ce se vor realiza. De asemenea, trebuie să se prevadă consecințele întârzierilor în finalizarea unor etape, ce se găsesc într-o măsură mai mică sub controlul cercetătorului, și să se aibă în vedere soluții de schimb.

În al doilea rând, apare ca o necesitate stringentă planificarea modului de gospodărire a mijloacelor materiale de care se dispune: cât poate fi cheltuit pentru documentare, pentru multiplicarea instrumentelor, pentru retribuirea personalului auxiliar, pentru prelucrarea datelor la calculator etc., astfel încât să nu se omită nici un amănunt - mergînd pînă la cheltuieli pentru creioane și radiere - și finalmente să ne putem încadra în volumul fondurilor avute la dispoziție.

Apoi, se pune problema muncii în cadrul echipei de cercetare



tare; fiecare membru trebuie să cunoască exact ce are de făcut și ce responsabilități îi revin, atât din punctul de vedere al cercetării științifice, cât și din cel administrativ-organizatoric. Dacă pentru întregirea echipei este nevoie și de un corp de auxiliari (de exemplu, operatori de anchetă), atunci sarcinile pe această linie se măresc simțitor. În faza pregătitoare investigației se va urmări cu toată atenția problema selecției acestora, pe baza unor criterii sau probe stabilite, precum și instruirea lor în legătură cu munca pe care o vor avea de desfășurat.

Un alt set de aspecte ce se cer rezolvate cu maximum de atenție tot în faza pregătitoare este cel legat de condițiile de lucru și de viață pe teren. De regulă, este necesar și recomandabil să se ia legătura din vreme cu organele locale atât pentru a fi încunoștiințate despre obiectivele cercetării (solicitând, dacă este cazul, aprobarea desfășurării ei), cât și pentru a primi un sprijin prețios în privința cazării, a asigurării mesei etc.

Mai menționez, în sfârșit, necesitatea programării altor aspecte privind desfășurarea muncii propriu-zise în teren:

- ce sarcini zilnice revin unei persoane,
- cum și la ce intervale se efectuează controlul operațiilor,
- locul și datele unor eventuale ședințe în plen ale echipei,
- când se efectuează predarea informațiilor culese etc.

Iată doar câteva din noianul de probleme organizatorice ce se ridică pentru orice cercetare sociologică de teren de mai mare amploare. Se observă, mă, în mare, aceste aspecte pot fi grupate în două categorii: unele legate de activitatea de cercetare științifică propriu-zisă și altele de asigurarea condițiilor pentru buna ei desfășurare. Din acest motiv, pe lângă un "comandament" științific, care să coordoneze primele aspecte, e nevoie și de una sau mai multe persoane cu atribuții strict organizatorice.

#### 1.10. Studiul pilot

În cazul cercetărilor sociologice ce vizează probleme sau utilizează instrumente de studiu inedite, intervine, de regulă, o etapă ce se poate plasa pe o poziție intermediară între pregătirea



cercetării și investigația propriu-zisă, numită studiu pilot. Este, în fapt, un fel de "repetiție generală", a cărei nevoie se simte mai cu seamă în cazul anchetelor. Sintetizând la maximum funcțiile studiului pilot, putem afirma că acest moment are următoarele obiective principale:

- a) testarea instrumentelor de culegere a datelor (în mod cu totul deosebit a chestionarelor);
- b) familiarizarea operatorilor cu instrumentele și, în general, cu munca de teren;
- c) obținerea de sugestii în vederea prelucrării datelor

Dacă e vorba de o anchetă, se va alege un eșantion redus de persoane - de la câteva zeci la 100 - , a cărui reprezentativitate nu ne interesează decât în măsura de a putea surprinde în cadrul său situații cât mai diverse, eșantion care nu se suprapune, de regulă, cu cel ales pentru a fi efectiv investigat. Acestor persoane din minieșantionul de probă li se aplică chestionarul, în condiții cât mai apropiate de cele ale cercetării. Se vor obține astfel, cum spuneam, informații legate de felul în care "funcționează" instrumentele de cercetare, se vor depista întrebările cu formulări defectuoase sau cele la care nu s-au prevăzut corect variantele de răspuns. De asemenea, vor apărea informații asupra atitudinii subiecților față de cercetare și vor fi puse în evidență acele întrebări "dificile" ce pot produce distorsionarea răspunsurilor la întreg chestionarul și multe, foarte multe alte aspecte capabile să amelioreze forma de prezentare și conținutul întrebărilor.

Asupra celui de al doilea aspect, pomenit la punctul b), iarăși nu cred că este nevoie să insist prea mult. Oricine a lucrat cu operatori de anchetă amatori știe câte probleme se ridică la completarea primelor chestionare, câte greșeli apar datorită faptului că în pregătirea lor "teoretică" pentru cercetare nu se pot prevedea toate situațiile concrete ce apar pe teren. Studiul pilot îi va ajuta să cunoască mai bine structura chestionarului, să interpreteze în mod uniform conținutul întrebărilor, să se deprindă cu tehnica de înregistrare (codificare), să găsească formulele cele mai potrivite de deschidere a discuției cu subiecții ș.a.m.d.



În sfîrșit, datele obținute pe acest mic eșantion pot furniza sociologului o indicație prealabilă asupra tendințelor ce se vor manifesta în populația de investigat, fapt care-l va ajuta să-și planifice mai eficient programul de prelucrare finală a informației, sugerîndu-i ipoteze suplimentare de lucru.

Se înțelege că în urma studiului pilot va avea loc definitivarea chestionarului și o reinstruire a operatorilor, axată pe problemele ce au apărut. În plus, acest studiu va permite să se evalueze mai exact timpul necesar realizării anchetei și sarcinile ce le poate îndeplini un operator pe parcursul unei zile de lucru. Astfel, sînt posibile corecții în programul de cercetare pe teren.

Studiul pilot, în aparență, mărește costul unei anchete prin faptul că introduce operații suplimentare. În fapt însă, aceste pierderi sînt înzecit compensate de cîștigurile și avantajele care-i urmează. Într-adevăr, odată operatorii dispersați în teren, orice schimbare în tactica și strategia inițiale devine practic imposibilă. Un chestionar tipărit și pus în lucru nu poate fi schimbat pe parcurs decît cu mare greutate și cu multe riscuri. Chiar dacă există posibilitatea reunirii operatorilor după prima zi de anchetă, pentru analiza problemelor practice ivite, nu există garanța că eventualele corecturi în chestionar sînt efectuate și respectate în continuare de toți operatorii. Dacă dispersarea în teren este ridicată, astfel de reuniuni nu sînt posibile și toate întrebările defectuoase vor trebuie pînă în urmă sacrificate; același lucru se va întîmpla și atunci cînd se lucrează prin auto-completarea chestionarelor, distribuite simultan întregului eșantion.

Pe de altă parte, dacă anumiți operatori de anchetă - cînd se folosește această tehnică - nu înțeleg exact sensul unor întrebări, ei vor introduce erori sistematice a căror depistare este, evident, preferabilă înaintea cercetării propriu-zise, cînd pot fi eliminați și operatorii ce se dovedesc incapabili să desfășoare o muncă de calitate. Chiar dacă avem posibilitatea să verificăm pe teren activitatea operatorilor, pînă cînd sînt supuși toți unui control serios, o parte a cercetării este deja epuizată iar erorile introduse și eventualele stereotipuri dobîndite cu greu se mai pot elimina în continuare.



Iată suficiente motive, care, deși prezentate pe scurt și fără pretenția de exhaustivitate, pledează pentru respectarea acestei sugestii metodologice (prezentă mereu în manuale, dar rareori pusă în practică) conform căreia majoritatea cercetărilor sociologice au nevoie de o probă prealabilă, numită studiu pilot.

## 2. Valorificarea informației

Dacă despre cea de a doua etapă mare a cercetării nu am găsit de cuvîntă să spun nimic în acest capitol introductiv și general, lucrul se datorează desigur aspectelor cu totul specifice fiecărei metode și tehnici de cercetare, astfel încît însăși tratarea acestora se va suprapune în bună măsură cu prezentarea activităților de culegere a informației. Voi trece deci la ultima etapă, care nici ea nu se pretează la deosebit de multe analize de ordin metodic cu caracter de generalitate.

a) După terminarea fazei de culegere a datelor, cercetătorul sau echipa de cercetare trebuie să procedeze la o verificare a acestora, din punctul de vedere al completitudinii, exactității, al modului în care sînt trecute în formularele de înregistrare, al acurateței scrisului etc. E nevoie de acest lucru atît pentru a depista unele lipsuri și erori strecurate în procesul de investigație, cît și pentru a preveni eventualele erori ulterioare datorate unor neuniformități de scriere, ilizibilități ș.a.m.d.

b) Următorul pas îl constituie prelucrarea sumară a datelor. Așa cum am spus în capitolul anterior, în marea lor majoritate, datele sociologice sînt de tip statistic și atunci această prelucrare sumară înseamnă construcția unor tabele simple de frecvențe (tabelarea) însoțită de procente, medii, indicatori de dispersie etc. Aceste tabele ne vor arăta, într-o primă formă, felul în care se distribuie indivizii în clasele caracteristicilor construite, tendința lor de grupare sau împrăștiere în jurul valorilor mijlocii.

E de la sine înțeles că de multe ori este nevoie de o fază prealabilă, pregătitoare acestei prelucrări: postcodificări, analize de conținut a unor texte sau întrebări deschise din chestionar etc. Dacă prelucrarea datelor se realizează pe calculator, trebuie ca informația să fie adusă la o formă adecvată pentru a putea fi transferată în memoria acestuia; uneori este nevoie de



transcrierea ei pe imprimare speciale, alteleori poate fi transpusă pe cartelele perforate direct de pe instrument (chestionar, formular de înregistrare etc.)

1) Urmează apoi un alt pas important, și anume prelucrarea și analiza datelor în profunzime, adică îmbinarea variabilelor (corelații, asocieri, regresie) și construirea de modele de dependență. Rostul acestei analize îl constituie testarea ipotezelor explicative avansate la începutul cercetării și obținerea unor informații de natură predictivă asupra fenomenelor sociale. Operațiile de finețe se realizează, chiar și pe calculator, după analiza sumară, întrucât vor trebui luate în considerare aici doar variabilele ce s-au dovedit relevante, capabile să dea o descriere adecvată fenomenelor reale reflectate-

2) Pe parcursul tuturor fazelor de prelucrare a datelor și apoi după terminarea lor, se vor degaja concluziile investigației, fie sub forma unor elemente pur descriptive, fie sub formă explicativă sau predictivă. Toate acestea vor fi reținute, clasificate, analizate, puse în legătură cu schema explicativă prealabilă, cu ipotezele care au ghidat cercetarea.

Ultimul act al oricărei investigații îl constituie redactarea raportului de cercetare. Folosesc aici acest termen într-un sens mai general decât cel obișnuit. De regulă, prin raport de cercetare se înțelege materialul întocmit de sociolog pentru beneficiarul care a finanțat cercetarea. Prin extensie, expresia poate fi utilizată și dacă rezultatele se fructifică sub formă de articol, comunicare, studiu de culegere sau chiar carte. Desigur că fiecare dintre formele respective implică exigențe specifice în redactarea și prezentarea materialului. Într-o carte, de exemplu, se vor accentua incomparabil mai mult aspectele teoretice ale problemei, se vor face incursiuni mai vaste în istoricul cercetării sale și, evident, se vor da mult mai multe detalii decât într-un articol sau comunicare. Dacă materialul este destinat unui beneficiar, el va trebui să aibă o prezentare grafică deseori de sugestivă, concluziile să fie clare și exacte, și în general, textul să fie limpede și concis. Tot aici se vor accentua cu precădere laturile aplicative ale studiului și soluțiile desprinse din cercetare, în scopul ameliorării aspectelor sociale urmărite.



Dar cred că este mai bine să schițez, după Th. Caplow (1970), criteriile ce pot fi folosite în evaluarea unui raport de cercetare. Ele vizează patru puncte esențiale:

a) Enunțul problemei studiate. Problema de studiu trebuie enunțată cu claritate în raport, astfel încât să nu existe ambiguități iar, dacă e posibil, să se prezinte ipotezele formalizate și metodele de verificare pentru acestea.

Un studiu sociologic se cuvine să poată fi evaluat și prin semnificația problemei cercetate, mergând de la o valoare maximă, când soluțiile interesează întreaga masă de sociologi sau un public foarte larg, până la una minimă, când soluția este absurdă, imposibilă sau fără valoare.

În al treilea rând, un studiu se poate caracteriza prin gradul de cuprindere a documentației existente în problema respectivă. Aprecierea studiului este aici, cum spuneam, în funcție de destinația sa, deci nu se poate pretinde întotdeauna documentația să apară cât mai complet. (E vorba de textul redactat și nu de munca de pregătire a cercetătorului!)

b) Descrierea tehnicilor de cercetare. În esență, e vorba la acest punct ca din textul raportului să reiasă faptul că:

(i) metodele, tehnicile și instrumentele folosite sînt adaptate problemei;

(ii) eșantionul, în cazul cînd se folosește, are un grad de reprezentativitate satisfăcător, cu precizarea limitelor de confidență pentru valorile calculate și a nivelului de probabilitate cu care diferiți coeficienți sînt semnificativi;

(iii) metodica cercetării poate fi reluată de către alți cercetători în scopul verificării concluziilor obținute, de unde rezultă că este nevoie de o prezentare suficient de detaliată a acesteia.

c) Prezentarea concluziilor. Indiferent de natura raportului de cercetare, concluziile trebuie prezentate în mod clar și precis, doar gradul de detaliere a lor fiind în funcție de destinația materialului scris. O prezentare a concluziilor poate fi considerată clară dacă un cititor cu formație medie este capabil să înțeleagă rezultatele sale. Nu de puține ori sociologii folosesc un limbaj prețios, care nu are finalmente alt rol decît să ascundă



slăbiciunile de conținut.

d) Interpretarea rezultatelor. Desigur, e vorba aici de faptul, că în întreg discursul argumentația trebuie să fie corectă și să nu se ascundă, prin tot felul de artificii, punctele slabe ale ei. Se poate da și o evaluare a importanței și utilității rezultatelor obținute, din punct de vedere științific sau din cel practic aplicativ.

În sfârșit, termin acest capitol tot cu o sugestie a lui Caplow. Marea masă de documente ce rămâne în urma unui studiu sociologic nu trebuie nici ea neglijată. Aceste documente (chestionare, listinguri de calculator, tabele manuale etc.) se cer clasificate și păstrate în arhivă pentru situații în care alți cercetători ar dori să refacă concluziile investigației, cu aceleași proceduri de analiză sau cu altele.



### Cap. III. CALITATIV SI CANTITATIV IN STUDIUL FENOMENELOR SOCIALE

Discuțiile asupra raportului calitativ-cantitativ în abordarea realităților sociale ocupă numeroase pagini, atât în tratatele generale de sociologie, în lucrările de filosofie socială, cât și în studiile cu caracter strict metodologic sau metodic-tehnic. Cu siguranță că problema este importantă din toate punctele de vedere, începând cu cel general filosofic și terminând cu cel practic, legat de tehnicile investigațiilor empirice. Pe de altă parte, este ușor, de închipuit că amplitudinea unei asemenea dezbateri ca cea stîrnită de relația cantitativ-calitativ este rezultatul direct al unei diversități enorme de poziții, mergînd de la cele care susțin imposibilitatea cuantificării fenomenelor sociale și pînă la cele care văd în cuantificare și măsurare singurele operații capabile de a scoate aceste științe, în speță sociologie, din starea de subdezvoltare în care, pe drept sau pe nedrept, se consideră adesea că ele se află.

În rîndurile care urmează, n-o să angajez o dezbatere cu caracter general asupra problemei enunțate în titlu, care n-ar face decît să mai adauge cîteva opinii - mai mult sau mai puțin interesante - la multitudinea celor care există deja. Voi încerca să precizez unele concepte de bază, privite în special prin prisma utilizării lor în cercetările concrete, ajutînd, sper, prin aceasta la o înțelegere mai clară a problematicii, fapt care va putea eventual constitui punct de plecare pentru unele generalizări sau discuții epistemologice-metodologice.

Nu îmi ascund totuși poziția de susținător al introducerii procedurilor cantitative în sociologie, considerînd că, în momentul de față, este nevoie de a sprijini din toate puterile orientările angajate pe calea aplicării în științele sociale a metodelor validate de științele pozitive. Asistăm acum, probabil ca o reacție la excesele de empirism și de aplicare a matematicilor fără discernămint, la o recrudescență a concepțiilor privind singularitatea metodologică a științelor sociale, la o repliere a unor cercetători în sfera metodelor subiective, la atacuri virulente la adresa



încercărilor de cuantificare și măsurare. Mă voi strădui să arăt că aspectele cantitative au un câmp mult mai larg de manifestare în viața socială decât se crede în general, că noțiunea de cantitativ nu este identică cu, ci mai largă decât cea de măsurare, că prin transferul de la anumite niveluri la altele ceea ce nu este măsurabil devine măsurabil, deci că o serie de critici sînt, în esența lor, viciate de o înțelegere mult prea îngustă a obiectului asupra căruia se exercită.

### 1. Precizarea noțiunilor

Fără a intra în probleme epistemologice de mare subtilitate va fi suficient, pentru cele ce ne interesează aici, să observăm că orice știință empirică studiază un domeniu al realității, deci un univers de elemente a căror existență este presupusă a fi independentă de acțiunea subiectului cercetător, realizînd în realitatea empirică un dublu decupaj: cel referitor la obiectele materiale (în sensul larg al termenului) și cel care privește o anumită gamă de caracteristici ale acestora. Obiectele materiale le voi numi obiecte-suport, spre a le distinge de cele științifice, construite de cercetător, și le voi considera ca purătoare a unor însușiri, înglobînd în această noțiune și cele de trăsătură, proprietate, caracteristică etc. Este clar că științe diferite pot porni la studierea aceluiași obiecte, dar privesc din perspectiva unor însușiri deosebite. Așa de exemplu, și fizica și chimia se ocupă de studiul atomilor, dar fiecare dintr-o altă perspectivă. În sociologie, asemenea obiecte pot fi indivizii umani, grupuri - mai mari sau mai mici - instituții, unități teritoriale sau de habitat etc.

Din cele de mai sus rezultă imediat că nu este niciodată suficient, pentru a desemna obiectul unei investigații, să se indice obiectul material vizat, ci trebuie adăugate și însușirile cercetate. Astfel, nu e suficient să spunem că studiem, de pildă, un microgrup, ci va trebui să precizăm și perspectiva din care-l privim; ne pot interesa relațiile interpersonale în cadrul său, frecvența cu care este exprimată o anumită opinie, atitudine etc. de membrii grupului, rețeaua prin care circulă informațiile de o anumită



mită natură între indivizi și multe alte probleme. Tot așa, nu ne va interesa satul sau un sat în sine, ci ca o comunitate caracterizată de fenomene specifice ce se petrec în interiorul său ori în raporturile cu alte unități. Nici măcar în studiile cu caracter monografic nu se studiază unitatea monografiată în sine, ci prin prisma unui complex de manifestări ale ei, set care, de astă dată, se urmărește a fi cât mai complet posibil. Așadar, în ultima instanță, corespondentul real al obiectului științific îl constituie aceste trăsături și relații, caracterizând obiectele materiale suport.

O primă - și cea mai frecventă - accepțiune a termenilor de calitativ și cantitativ rezultă din modalitatea de abordare, de investigare și interpretare a însușirilor și relațiilor respective. Intr-adevăr, aceste trăsături se manifest sub forma unor stări ce pot fi percepute fie direct cu organele de simț, fie prin intermediul unor aparate de genul celor din fizică, chimie, etc., fie doar mental, mai ales în cazul fenomenelor sociale, prin încadrarea lor în niște grile socialmente constituite. Din punctul de vedere al culorii, un obiect poate fi perceput ca verde, albastru etc., din alt punct de vedere el poate fi categorisit ca mare sau mic, din altul poate fi apreciat drept urât sau frumos etc. În studiile sociologice, găsim nenumărate exemple prin care să punem în evidență stările unei însușiri. De pildă, o persoană (obiectul), avînd statutul de student (însușirea) poate învăța la facultatea de filologie, istorie, matematică etc. (stările); un om al muncii (însușirea) poate lucra în domeniul agriculturii, industriei, comerțului etc. (stările) ș.a.m.d.

Exemplele de mai sus sînt ilustrări tipice de însușiri evaluate din punct de vedere calitativ, fiecare stare a însușirii fiind deci o calitate a acesteia. Abordarea sau evaluarea cantitativă este posibilă, în principiu, atunci cînd stările trăsăturii posedă un caracter de gradualitate, cînd trecerea de la una la alta înseamnă o creștere sau descreștere a intensității manifestării. Această variație poate fi continuă sau discretă. În primul caz, trecerile de la o stare la alta sînt line, diferențele dintre poziții diferite dar apropiate apar greu de sesizat, deci delimitarea exactă a unor stări se face cu dificultate, practic orice



"tăietură" pe scala de variație presupunând un oarecare grad de convenționalitate; din contră, în cazul secund, trecerile sînt bruste, diferențele dintre stări foarte clare și nete. Multe trăsături ale indivizilor umani, mai ales de natură psiho-socială, deși nu avem mijloace clare de a evidenția acest lucru, sînt considerate a avea o manifestare sub formă continuă. Astfel, gradul de satisfacție al omului în raport cu ceva este deseori apreciat a avea un interval de variație continuă, de la o stare de totală insatisfacție la una de maximum. Sau, în sociologia politică, nu o dată convingerile politice ale oamenilor sînt plasate (imaginar vorbind) pe o scală ce ar merge de la extrema dreaptă la cea stîngă. Chiar dacă practic, în astfel de cazuri, sînt evidențiate doar unele stări discrete - cîteva grade de satisfacție: mare, mică, moderată etc. - de exemplu - ele nu sînt considerate decît niște substitute, aproximări foarte imperfecte ale unei scale obiective, capabilă să evidențieze o infinitate de nuanțe.

Evaluarea cantitativă a unor asemenea însușiri poartă denumirea de măsurare, al cărei rezultat - măsura - este o expresie numerică a intensității însușirii studiate. Nu o dată măsurarea este definită drept operația prin care se atagează diferitelor stări ale fenomenului numere reale. Definiția aceasta curentă mi se pare totuși prea largă, deoarece simpla atașare de numere nu înseamnă totdeauna o măsurare, în sensul consacrat pe care-l are acest termen. De exemplu, dacă întocmim o clasificare a studenților din anul II filosofie după cursul de limbă străină pe care-l urmează, obținem 6 categorii ce pot fi numerotate astfel:

1. Engleză
2. Franceză
3. Germană
4. Italiană
5. Rusă
6. Spaniolă

Ulterior, lucrînd cu grupele respective, ele pot fi indicate doar prin simbolurile numerice acrodate, găsindu-ne deci într-o situație care se încadrează exact în definiția amintită. A spune însă că numerele atagate reprezintă o măsură înseamnă a goli termenul de măsură de orice conținut.



Prin urmare, măsurarea trebuie să constituie o operație mult mai complexă decât simpla atribuire de numere. În esență valorile acordate trebuie să se coreleze, prin mărimea lor, cu nivelul sau intensitatea însușirii evaluate. Desigur că o regulă formală, capabilă să surprindă exact semnificația corelației în cauză, este greu de dat, ceea ce nu înseamnă că, în practică, nu vom putea distinge, de cele mai multe ori, simplele atașări arbitrare de numere de cele în care valorile evaluează cantitativ însușirea. Dacă totuși dorim o definiție, vom putea, provizoriu și în mod suficient pentru necesitățile cursului de față, să pretindem că, pentru a putea vorbi de măsurare, în mulțimea valorilor atribuite să fie posibilă realizarea cel puțin a unei operații aritmetice simple - scăderea - operația și rezultatul ei având o semnificație și un sens. Mai clar spus, diferențele aritmetice dintre valorile numerice trebuie să reflecte diferențele reale între gradele de manifestare a fenomenului.

Un astfel de lucru nu este desigur posibil în situații ca cea imaginată mai sus. Numerele atribuite n-au nici măcar o semnificație ordinală (căci scrierea putea urma foarte bine și altă ordine decât cea alfabetică) și cu atât mai puțin una care să ne permită să le utilizăm în vreo operație matematică.

Realizarea în practică a operației de măsurare presupune rezolvarea următoarelor probleme;

a) Delimitarea printr-o anumită procedură a universului obiectelor suport (spre a ști pe cine măsurăm)

b) Alegerea trăsăturii, însușirii etc. ce trebuie sopușă operației de măsurare (ce măsurăm)

c) Alegerea unei unități de măsură sau etalon (cu ce măsurăm)

d) Definirea procedurilor concrete de aplicare a etalonului sau, altfel spus, stabilirea regulilor de atribuire a numerelor pentru obținerea măsurii (cum măsurăm).

Cele mai mari dificultăți de măsurare în sociologie sînt legate de ultimele două puncte mai sus enunțate. În științele naturii și în viața cotidiană, pentru efectuarea celor mai multe măsurători există unități naturale, fizice de măsurat, iar realizarea



operației propriu-zise este tot una fizică, fie directă prin manipularea etalonului, fie indirectă prin intermediul aparatelor. Determinarea masei, a lungimii, a intensității sau tensiunii curentului electric etc. sînt cîteva exemple simple de măsurători cu unități direct manipulabile (metrul, kilogramul) sau indirect (amperul, voltul). Măsurarea intensității unei atitudini sau a gradului de adeziune la o normă socială etc. sînt cazuri tipice din științele umane în care nu avem nici unități materiale și nici instrumente bine standardizate pe care să putem citi valorile ca la voltmetru sau ampermetru. Atît etalonul cît și instrumentul (teste, chestionare) trebuie elaborate pe baza unui suport teoretic - ce nu este nici el prea solid - și pe cea a unei experiențe practice în studiul fenomenelor sociale.

De cele mai multe ori în sociologie - și nu numai aici - pentru a ajunge să măsurăm o însușire, trebuie să trecem printr-o etapă intermediară ce poartă numele de cuantificare. Măsurarea și cuantificarea nu sînt deci două noțiuni identice, cum adesea se mai prezintă; prima desemnează operația concretă prin care se obține măsura, pe cînd a doua ansamblul de operații de natură conceptuală prin care se face posibilă măsurarea. Concret, în cazul sociologiei, cuantificarea presupune o analiză aprofundată a fenomenului de măsurat, descompunerea conceptului în dimensiuni, determinarea indicatorilor, construcția indicilor și, în fine, agregarea acestora într-un indice sintetic ale cărui valori reprezintă măsura. Deci cuantificarea este o operație efectuată la nivelul teoretic și metodologic, implicînd, în esență, punerea în evidență, dezvoltarea, descoperirea - în multitudinea de determinații pe care le au fenomenele sociale - a aspectului cantitativ și indicarea formelor concrete sub care el se manifestă și poate fi perceput. Deci tot cuantificarea este cea care duce la alegerea unității de măsură și la elaborarea scalei pe care se găsesc valorile - rezultat al măsurătorii. În fond, orice măsurare e precedată de cuantificare, numai că în cazul trăsăturilor foarte simple acest lucru nu este vizibil, intrucît efortul teoretic și metodologic pentru prepararea măsurătorii este minim.

In concluzie, voi afirma că aspectul cantitativ în evaluarea manifestărilor unei însușiri, trăsături etc. apare atunci cînd



este posibilă operația de măsurare, în sensul mai sus prezentat. În fapt, aceasta este și cea mai frecventată accepțiune dată cantitativului și nu de puține ori este și singura despre care se vorbește în literatura de specialitate. Din păcate, confuzia între cantitativ și măsurare nu este deloc benignă, căci o serie de atacuri la adresa cantitativismului se hrănesc din argumente variabile doar într-o eventuală confruntare cu abuzul de măsurare.

Intr-adevăr, abordarea cantitativă și măsurarea nu se suprapun pentru simplul motiv că prima face apel și la alte operații și tehnici decât cele metrice. Este vorba, în speță, de operația de numărare. Nimeni nu gîndește probabil să excludă numărarea dintre componentele cantitativului, dar, curios, foarte puțini sînt cei care fac distincția elementară dintre numărare și măsurare. Chiar și un autor care se ocupă într-o carte întreagă despre problema măsurării (l-am numit pe I. Mărginean, 1982) confundă cele două chestiuni, lăsînt să se înțeleagă faptul că măsurarea ar include și numărarea. Dar noi putem înțelege acum că în timp ce prima operație face apel, așa cum s-a văzut, la manipularea însușirilor obiectelor, a doua (numărarea) operează cu obiectele însele. Nu măsurăm oameni, ci înălțimea sau greutatea lor, în schimb efectiv numărăm oameni sau alte entități materiale purtătoare de însușiri.

Așadar, putem vorbi de abordare cantitativă la două niveluri: al însușirii (măsurarea) și al obiectului (numărarea). Inversînt perspectiva, atunci și calitativul va apare, prin simetrie, în dublă ipostază: un gen de abordare la nivelul însușirii și altul la cel al obiectelor-suport. Dacă în primul caz termenul se menține în accepțiunea sa obișnuită, cunoscută și larg folosită, în al doilea, el pare mai greu de utilizat și probabil ar trebui găsit unul mai adecvat. Dar cum importantă nu este inventarea de cuvinte, voi menține tot expresia de calitativ și aici, predizînd doar sensul în care o folosesc.

Va trebui deci să vedem care este rostul numărării, dacă și cînd intervine ea în studiile științifice și care este specificul acelor științe care fac apel la numărare. Intr-adevăr, problemele acestea nu sînt lipsite de sens deoarece știm bine că ramuri întregi ale unor științe nu utilizează operația de numărare, decît poate în mod accidental și pentru chestiuni marginale. Astfel pen-



tru a studia structura atomului de hidrogen, de exemplu, este suficient să se studieze un atom iar rezultatele obținute sînt imediat transferate asupra întregii mulțimi de atomi de hidrogen din univers: la fel pentru molecule, sau substanțe complexe. Proprietățile simple ale unor elemente, din universul fizic mai ales, ne dau această posibilitate de a face abstracție de obiectul material purtător de însușiri și de a generaliza o anumită proprietate, relație sau structură la toate obiectele de aceeași specie.

Uneori însă astfel de generalizări nu sînt posibile; însușirile prezintă variații mari de la individ la individ, raporturile și conexiunile depind de purtătorii materiali ai elementelor ce se combină. Atunci va fi nevoie să numărăm; să vedem cîți indivizi au sau, mai exact, ce proporție din rîndul lor are un comportament sau altul, ce proporție se află într-o stare sau alta etc. Legitățile detectate în astfel de situații vor fi de tip probabilist iar în abordarea efectivă a realității se va face apel la instrumentul de analiză numit statistică. Nu este greu să observăm că aceasta este situația pentru majoritatea fenomenelor sociale; nu este greu deci să înțelegem de ce științele sociale fac apel în mod curent la statistică; nu este greu să intuim că în studiul fenomenelor sociale aspectul cantitativ legat de numărare nu este unul secundar, neglijabil, ci el are în realitate tot atîta importanță ca și cel legat de măsurare.

Din cele spuse pînă aici, rezultă deci, că pe de o parte, însușirile obiectelor pot fi descrise fie de o manieră calitativă, fie de una cantitativă. Pe de altă parte, o însușire poate fi urmărită, în formele sale de manifestare, fie prin inventarierea obiectelor suport - adică încercînd să se determine ce fracțiuni din "universul" de obiecte studiat posedă cutare sau cutare formă de manifestare a caracteristicii - fie, prin analiza unui singur obiect, care este capabil să reprezinte corect o întreagă clasă de obiecte identice. Vom avea deci patru tipuri fundamentale de metode de analiză, considerînd simultan însușirea (primul termen) și obiectul (al doilea):

- a) Metode calitativ-calitative
- b) Metode cantitativ-calitative
- c) Metode calitativ-cantitative
- d) Metode cantitativ-cantitative



*col-col.*  
Primul gen cuprinde metodele care abordează trăsătura din punct de vedere calitativ și nu ia în considerare întinderea sau frecvența cu care se realizează diferitele stări ale trăsăturii respective. Al doilea tip încearcă să măsoare intensitatea manifestărilor, iarăși făcând abstracție de obiectul concret. Al treilea este similar primului și al patrulea celui de al doilea, introducând, în plus, fiecare și obiectul ca entitate ce contează în analiză.

Aceste două din urmă metode sînt deci cele care fac apel la statistică, iar distincția dintre ele revine la distincția ce se face în statistică între procedurile de analiză calitativă, și cantitativă sau nonparametrice și parametrice.

Abordarea cantitativă din unghiul obiectului este impusă de faptul că fenomenul studiat nu posedă, așa cu, spuneam, un grad de invarianță suficient de ridicat, în raport cu diferite condiții, astfel încît el să se mențină constant de la un individ la altul. Dacă știm, de exemplu, că un om este "muncitor", obținem o informație (suplimentară față de cea conținută în definiție) mult mai slabă decît dacă spunem că un atom este de oxigen. Despre atomii de oxigen se poate afirma că toți au masa atomică 16, despre muncitori că au în medie o retribuție de 2500 lei (de pildă). Pentru formarea moleculei de apă, un atom de oxigen intră întotdeauna în combinație cu doi atomi de hidrogen; muncitorul, pentru întemeierea unei familii, își alege, cel mai adesea, un partener dintr-un mediu social apropiat etc. Iată tipuri de aserțiuni care se folosesc într-un caz sau celălalt.

Abordarea nestatistică se întâlnește deci în foarte multe domenii ale științelor naturii, deși chiar și cele mai dezvoltate dintre acestea fac apel la instrumentul statistic. Ea se folosește curent și în științele socio-umane, mai cu seamă atunci cînd, neavînd de a face cu o identitate a obiectelor, se caută pentru analiză unul tipic sau reprezentativ pentru o anumită clasă de obiecte. Așa, de pildă, studiind capitalismul în Anglia, Marx a abordat un caz tipic sau clasic sau un model de sistem social capitalist.

Între această din urmă formă de analiză și cea statistică nu este o linie de demarcație riguros trasată. Alegerea <sup>cazului</sup> tipic (tipice) se poate baza pe o analiză prealabilă a





masei de unități, tot așa cum un studiu statistic poate urma unuia de caz. Această îmbinare de metode se practică în mod curent și cu succes în cadrul științelor socio-umane, putând duce la rezultate foarte bune. Analiza de caz permite studierea fenomenelor în adâncimea lor, descrierea și explicarea conexiunilor între diferite însușiri ale obiectului, urmărirea în timp a manifestărilor acestora etc. Analiza statistică a unei populații desprinde ceea ce este general, repetabil, măsoară frecvențele cu care apar anumite stări, determină tendințele de evoluție, stabilește regularitățile care se observă numai în cazul unei mulțimi de indivizi (fenomenele de masă). Așadar, nu numai că cele două tipuri de metode nu se exclud, nu se opun unul altuia, cum se mai susține uneori, dar ele se dovedesc a fi complementare, susținându-se reciproc în investigațiile empirice.

## 2. Scale de evaluare

Reluând acum problematica din unghiul de vedere al trăsăturii, voi încerca să schițez sumar care sînt instrumentele cele mai des utilizate pentru evidențierea formelor de manifestare ale acesteia, stărilor în care se poate găsi sau nivelului de intensitate manifestat. Ajungem astfel la ceea ce în literatura metodologică poartă denumirea de scale de măsură, termen care mi se pare inadecvent, pentru motive care vor fi înțelese imediat, motive legate de utilizarea abuzivă a cuvîntului "măsură"; prefer, de aceea, denumirea de scale de evaluare. Tipurile fundamentale de scală sînt în număr de patru - în cadrul lor putîndu-se găsi și unele subtipuri specifice-, mergînd de la o evaluare pur calitativă spre una pur cantitativă, în ordinea în care vor fi prezentate mai jos.

Scala nominală. Este instrumentul folosit în realizarea oricărei clasificări. Scala nominală evidențiază direct stările însușirii urmărite, stări care formează treptele sau elementele scalei. Cea mai simplă scală nominală este cea cu două trepte, scala dihotomică. De exemplu, pentru însușirea "sex", există două stări. "masculin" și "feminin", care constituie elementele scalei respective. Pentru însușirea "ocupație", putem construi o scală cu următoarele elemente: "agricultor", "muncitor", "tehnician", "intelectual", "funcționar", "alte ocupații".



Este evident că elementele scalei nominale, spre a fi identificate, poartă anumite etichete, sînt marcate de unele simboluri. Acestea, ca în exemplele de mai sus, pot fi cînvinte prin care să se sugereze conținutul noțiunii prin care se constituie o treaptă a scalei sau pur și simplu niște simboluri abstracte, fără legătură cu acest conținut, cum ar fi litere sau numere. Subliniez că eventuala folosire de numere pentru simbolizare nu ne dă dreptul să facem apel la nici o proprietate matematică a lor.

Așadar, elementele scalei nominale sînt stări pur calitative, distincte, în care un obiect sau altul se poate afla. Singurele cerințe formale ridicate de scala nominală sînt cele ale oricărei clasificări: orice individ să poată fi plasat pe o treaptă și numai pe una a scalei. Dacă abordarea la nivelul obiectului suport este cantitativă (statistică), atunci elementele scalei nominale devin clase în care se grupează indivizii, în funcție de însușirea studiată. Problema practică cea mai importantă este elaborarea unui criteriu de decizie cu ajutorul căruia fiecare individ să poată fi plasat în una și numai în una dintre clase.

Scala ordinală. Este un gen mai complex de scală decît cea precedentă, în sensul că posedă toate însușirile scalei nominale, avînd, în plus, proprietatea că între elementele ei se stabilește o relație de ordine. Spre deosebire deci de scala nominală, unde stările însușirii erau evidențiate prin elemente ale căror poziții în scriere erau indiferente, la cea ordinală treptele scării respectă o anumită ordine, reflectînd o ierarhie între stări, dată, îndeosebi, de intensitatea cu care însușirea evaluată se manifestă. Prin urmare, scala ordinală surprinde de o manieră calitativă o manifestare graduală a fenomenelor. Se pleacă cel mai adesea de la supoziția că fenomenul în sine are o variație cantitativă și că doar formele subtile de manifestare a lui, limitele cunoașterii noastre, imprecizia instrumentelor de înregistrare ne fac să nu-l putem surprinde în această manieră.

Dacă notăm elementele (treptele) scalei cu litere mari: A, B, C, ..., atunci relația dintre ele poate fi scrisă:



și se poate citi: "A mai mic decât B", "A se află înaintea lui B", "poziția A este inferioară poziției B", "starea A este mai puțin intensă decât B" ș.a.m.d.

Ca orice relație de ordine, și aceasta trebuie să îndeplinească unele condiții fundamentale:

a) Orice ar fi două elemente ale scalei, A și B, vom avea  $A < B$  sau  $B < A$ .

b) Intre două elemente diferite ale scalei A și B, nu putem avea simultan relațiile  $A < B$  și  $B < A$ . În cazul că s-ar întâmpla așa ceva, elementele A și B trebuie să coincidă.

c) Dacă  $A < B$  și  $B < C$ , atunci  $A < C$ . Este celebra proprietate de tranzitivitate a relației de ordine.

Exemple de scale ordinale folosite în sociologie se pot da foarte ușor. Multe întrebări în cadrul chestionarelor au variante de răspuns ce se constituie în elemente ale unor asemenea scale. De pildă, la întrebări de tipul: "Cum apreciați emisiunea...., prezentată în ziua... la televiziune?" apar, în chestionare, răspunsuri precodificate de forma:

1. Foarte bună
2. Bună
3. Mediocră
4. Slabă
5. Foarte slabă

formind o scală cu cinci trepte, în scopul de a evidenția opinia celui chestionat.

Adesea elementele scalei se simbolizează prin cifre. Aci, spre deosebire de cazul precedent, cifrele nu pot fi atașate, oricum, ele trebuind să respecte relația de ordine dintre stări. Putem numerota treptele scalei așa cum am făcut-o mai sus, în exemplul dat, ori putem să le numerotăm invers dar nu și în alt mod. Cu toate că scala ordinală încearcă să reflecte ceva din aspectul cantitativ, cu toate că numerele atașate treptelor reflectă ordinea acestora (deci facem apel la relația de ordine dintre numere) ele - numerele - nu pot fi interpretate în termeni de măsură deoarece nu reproduc nimic din aspectul cantitativ presupus; scala ordinală rămâne o scală de tip calitativ, chiar dacă reprezintă



tă, față de cea nominală, deja un pas spre abordarea cantitativă.

Scala de intervale. Este, în esență, rezultatul cuantificării fenomenului studiat, deci o scală de măsură. Elementele scalei se reprezintă în mod obligatoriu prin numere (reale), care, de data asta, nu mai joacă doar rolul de simboluri, ci reflectă efectiv magnitudinea manifestării însușirii. Specificul acestora scale, în raport cu următorul tip, îl reprezintă un grad ridicat de convenționalitate în construcție, în sensul că, pe lângă alegerea unității de măsură (a etalonului) - unitate care este totdeauna convențională -, cercetătorul fixează după bunul său plac și originea scalei. Mai exact spus, asemenea scale nu au un punct zero impus obligatoriu de manifestarea fenomenului.

Pentru a înțelege mai ușor specificul scalei de intervale, majoritatea manualelor dau ca exemplu, pe bună dreptate, o scală de măsură a temperaturii, de pildă scala Celsius. Aici punctul zero e ales convențional - temperatura de îngheț a apei -, el putând fi ales la orice alt nivel (și este efectiv ales la altul pe scara Fahrenheit).

Restricția fundamentală ce caracterizează scala de intervale este că două valori măsurate cu această scală nu se pot raporta una la cealaltă. Spre exemplu, dacă astăzi la ora 15 la București sînt  $30^{\circ}\text{C}$  iar la Brașov  $15^{\circ}\text{C}$ , nu este permis să se spună că la București aerul are o temperatură de două ori mai mare (sau că este de două ori mai cald) decît la Brașov. Că lucrurile stau așa o simțim fiecare dintre noi, și de obicei, nu facem astfel de greșeli de apreciere. Acum putem să și înțelegem lucrurile perfect, fiind suficient să ne imaginăm că punctul zero ar fi fost ales la nivelul de  $10^{\circ}\text{C}$ . Atunci, la București am fi avut o temperatură de 20 de grade iar la Brașov una de 5 grade și, prin urmare, dacă încălcăm restricția, ar trebui să spunem că la București e de patru ori mai cald ca la Brașov. Or, este absurd ca, în funcție de scală, același fenomen să poată fi atît de divers interpretat.

Comparații în termeni de rapoarte se pot face pe scala de intervale numai între diferențele dintre valori. Astfel, dacă în aceeași zi, la Cluj sînt  $20^{\circ}\text{C}$ , putem spune că diferența de temperatură dintre București și Cluj ( $30^{\circ}\text{C} - 20^{\circ}\text{C} = 10^{\circ}\text{C}$ ) este de



două ori mai mare decât diferența de temperatură între Cluj și Brașov ( $20^{\circ}\text{C} - 15^{\circ}\text{C} = 5^{\circ}\text{C}$ ). Unde am muta originea și oricum am schimba unitatea de măsură, acest raport rămâne constant. Astfel, punând, ca mai sus, zero în locul lui  $10^{\circ}\text{C}$ , la Cluj vom avea 10 grade. Pe noul termometru, diferența dintre București și Cluj va fi 10 grade ( $20 - 10$ ) iar între Cluj și Brașov 5 grade ( $10 - 5$ ) și raportul diferențelor se păstrează.

Această caracteristică a scalelor de intervale poate fi matematic precizată în felul următor: orice scală de intervale poate fi transpusă în alta printr-o transformare lineară.

Adică, dacă notăm cu  $x$  valorile pe care variabila le ia pe o scală, ele pot fi transformate în alte valori,  $y$  cu ajutorul formulei:

$$y = a + bx \quad (a \neq 0, \quad b \neq 0)$$

Coeficientul  $a$  ne arată cum se schimbă originea de la trecerea lui  $x$  la  $y$  ( $x = 0 \Rightarrow y = a$ ), iar coeficientul  $b$  cum se modifică unitatea de măsură (etalonul).

Fie acum două valori diferite ale lui  $x$ :  $x_1$  și  $x_2$ . Transformând scala, vom obține corespunzător două valori pentru  $y$ :

$$y_1 = a + bx_1$$

$$y_2 = a + bx_2$$

Se poate demonstra ușor că raportul  $x_1/x_2$  este diferit de  $y_1/y_2$ , adică:

$$\frac{a + bx_1}{a + bx_2} \neq \frac{x_1}{x_2}$$

Intr-adevăr, presupunând că cele două rapoarte ar fi egale, ar rezulta:

$$(a + bx_1) \cdot x_2 = (a + bx_2) \cdot x_1$$

sau

$$ax_2 + bx_1x_2 = ax_1 + bx_1x_2$$



deci

$$ax_2 = ax_1$$

și fiindcă  $a$  este diferit de zero:  $x_2 = x_1$

ceea ce este în contradicție cu presupunea că  $x_1$  e diferit de  $x_2$

În schimb, raportul diferențelor se conservă. Pentru demonstrație, să mai luăm o valoare distinctă  $x_3$ , careia îi corespunde:

$$y_3 = a + bx_3$$

Să calculăm diferențele:

$$y_2 - y_1 = a + bx_2 - a - bx_1 = b(x_2 - x_1)$$

$$y_3 - y_2 = a + bx_3 - a - bx_2 = b(x_3 - x_2)$$

Deci

$$\frac{y_3 - y_2}{y_2 - y_1} = \frac{b(x_3 - x_2)}{b(x_2 - x_1)} = \frac{x_3 - x_2}{x_2 - x_1}$$

ceea ce trebuia demonstrat.

Scala de rapoarte. Reprezintă un instrument de măsură mai puternic decât scala de intervale deoarece permite raportarea măsurilor obținute. Acest lucru rezultă din unicitatea punctului de origine (existența unui zero natural). De pildă, în măsurarea temperaturii o scală de rapoarte este scala Kelvin a temperaturilor "absolute". Singurul element convențional al scalei de rapoarte este unitatea de măsură. Formula de transformare a unei astfel de scale în alta de aceeași natură este:

$$y = ax$$

Formula reflectă tocmai faptul că printr-o astfel de transformare are loc schimbarea unității de măsură: la o unitate a lui  $x$  vor corespunde  $a$  unități pentru  $y$ , căci dacă  $x = 1$ , rezultă că  $y = a$ . Evident aici:



$$\frac{y_1}{y_2} = \frac{ax_1}{ax_2} = \frac{x_1}{x_2}$$

Exemple de astfel de nivele de măsură se pot da cel mai simplu din activitatea cotidiană. Lungimea unui corp, indiferent că e măsurată în metri, yarzi, coți etc. presupune o scală de rapoarte, căci zero corespunde mereu aceleiași situații: lipsa însușirii respective.

În domeniul sociologiei și al științelor sociale, în general, atât folosirea scalelor de intervale cât și a celor de rapoarte este mai puțin frecventă decât în cazul științelor naturii, unde însușirile se pretează mai ușor la măsurători. Totuși astfel de evaluări nu lipsesc nici în științele umane, ba așa putea zice că ele sînt mai numeroase decât mulți își închipuie. Într-adevăr, cînd se vorbește de astfel de scale, se au în vedere aproape exclusiv însușiri ale indivizilor umani cum ar fi inteligența, opinii, atitudini etc. În cazul unor asemenea însușiri este drept că puține se pot măsura cu adevărat. Există cîteva variabile simple ca vîrsta, retribuția, nivelul de școlarizare ș.a. ce pot fi măsurate relativ ușor, chiar pe scară de rapoarte. În rest, mai cu seamă pentru opinii și atitudini, întîlnim numeroase încercări de a construi scale de măsură, existînd deci o "tehnologie" specială - dezvoltată de psihologie și psihosociologie - avînd un asemenea obiectiv. Metode sofisticate, legate de nume celebre ca Guttman, Thurstone sau Likert, au fost imaginate spre a construi scale de intervale sau de rapoarte pentru a măsura asemenea însușiri. Nu mă voi ocupa de ele aici, căci am specificat de la început că voi accentua doar problematica specifică sociologiei, or, cum spuneam, aceste proceduri apar cu precădere în psihologia socială. Pe de altă parte, numai tratarea lor spre a fi corect înțelese și aplicate ar lua probabil un volum egal cu jumătate din cel al cursului de față.

Cînd spuneam că există și în sociologie mai multe posibilități de măsurare decât se întrevăd la o privire superficială, mă gîndeam la o altă chestiune și nu la aceste scale. Ar trebui observat că pentru sociolog unitățile de analiză nu sînt întotdeauna (și poate nici cel mai adesea) indivizii umani, ci grupurile



de indivizi, unitățile teritoriale, instituțiile etc. Or, cum voi arăta imediat, o serie de atribute ale acestora se lasă cu ușurință măsurate.

Pentru ilustrare, mă voi opri la un singur aspect care mi se pare esențial, nu numai pentru că pune în evidență varii posibilități de măsurare în sociologie, ci pentru că face o legătură directă între cele două aspecte ale cantitativului - măsurare și numărare - , aspecte pe care inițial le-am delimitat categoric, urmînd acum să arăt că este posibilă și o trecere de la unul la altul. Sintetic spus, ceea ce vreau să scot în evidență aici este constatarea că anumite proprietăți grupale (definite deci la nivelul grupului de indivizi) sînt măsurabile în virtutea faptului că ele au la bază numărarea în interiorul fiecărui grup, deci la un nivel inferior celui la care se realizează analiza.

Astfel, dacă unitatea de studiat este familia, putem defini o trăsătură a acesteia - s-o numim dimensiunea familiei - prin numărul de indivizi ce o compun. Vom obține astfel familii cu o persoană, cu 2, cu 3 etc. Aceste valori reprezintă evident o măsură a dimensiunii familiei și ele sînt elementele unei scale de rapoarte. Există un zero natural și se poate spune că o familie este de două sau de trei ori mai numeroasă decît alta.

Încă un exemplu, tot atît de simplu, poate fi următorul. Să presupunem că ne interesează grupele de studenți dintr-o universitate iar proprietatea studiată este promovabilitatea în sesiunea din vară. Cu ajutorul numărării, putem stabili, la nivelul fiecărei grupe, procentul studenților promovați în vară la toate examenele. Evident că aceste procente sînt o măsură a promovabilității în cadrul grupelor și că valorile obținute merg, în principiu, de la 0 (nici un promovat) la 100 (toți promovați), realizînd finalmente din nou o scală de rapoarte.

Această pistă deschisă, adică sesizată ideea de construcție a acestor scale, este clar că numărul lor se poate înmulți fără dificultate și este tot atît de limpede că nu rareori lucrăm cu scale foarte puternice fără să ne dăm seama de acest lucru. Ideea vine deci să sprijine o altă și anume că demersurile cantitative în abordarea fenomenelor sociale nu sînt deloc excepții. Numărarea, ca un prim nivel apoi măsurarea unor însușiri individuale și, în fine, măsurarea unor însușiri de grup, avînd la bază



numărarea, sînt tot atîtea ipostaze ale cantitativului și probabil nu singurele.

Indiferent de opiniile pe care teoreticienii le emit în legătură cu raportul cantitativ-calitativ, indiferent poate chiar de propria-i concepție, sociologul angajat în cercetarea empirică va avea mereu tendința de a utiliza scale de evaluare cît mai puternice de preferință scale de măsură, deoarece aceste instrumente de măsură au virtuți analitice superioare, îi deschid posibilități multiple de studiere a raporturilor dintre fenomene. De aceea apare uneori tendința - și atunci criticile calitativiştilor sînt pe deplin justificate și la obiect - de a "silui" fenomenele sociale, forțîndu-le să intre în cadrele strîmte ale unui instrument de măsură. Două puncte nevralgice mi se pare că trebuie relevate în acest sens:

- a) aplicarea scalelor ordinale acolo unde fenomenul nu manifestă o variație graduală;
- b) aplicarea scalelor de intervale și rapoarte în cazul în care variația graduală există dar măsurarea e inconsistentă, rezultat al unei cuantificări cu un grad excesiv de convenționalitate și al lipsei oricărui temei de a acorda valori manifestărilor fenomenului.

Această din urmă tendință poate fi ușor detectată în multe lucrări sociologice apărute în anii din urmă la noi. Fascinați de posibilitățile multiple de a "jongla" cu valori și de tennicile matematice complexe bazate pe scalele metrice, sociologii recurg la asemenea scale fără prea mult discernămint. O întrebare de opinie cu 5 sau 7 variante de răspuns ierarhice (de genul: foarte mult, mult etc.) e automat ransformată într-o scală de măsură, dîndu-se variantelor valori de la 1 la 5 ori de la 1 la 7, valori cu care se calculează apoi medii, indici de dispersie, coeficienți de corelație ș.a.m.d. Tentăția este ușor de înțeles, dar și în utilizarea noțiunii de măsură se cere .... măsură!

Revenind la problematica scalelor, să precizăm că acestora, ca de altfel oricărui instrument de observare a realității empirice, nu trebuie să le lipsească două proprietăți fundamentale: validitatea și fidelitatea.

O scală este validă dacă evalunază (măsoară) ceea ce



urmărește să măsoare. E vorba aici, în fond, de raportul dintre conținutul conceptului corespunzător însușirii și mijloacele de identificare construite. Dacă măsoarăm inteligența, să zicem, prin rezultatele pe care oamenii le-au avut în școală la matematică - ceea ce, de altfel, unora le poate părea destul de normal - atunci e aproape sigur că nu mi-am ales un instrument valid deoarece noțiunea de inteligență acoperă mult mai multe aspecte decât eventualele aptitudini pentru matematică. În ultimă instanță, determinarea validității revine la o analiză aprofundată a conceptelor și a modului de operaționalizare a lor. Exemplul cu inteligența este foarte sugestiv pentru că redă o caracteristică mai generală a multor noțiuni din domeniul științelor sociale: pluridimensionalitatea lor, bogăția de sensuri și semnificații. S-ar putea spune că practic orice concept din domeniul științelor noastre suferă o sărăcire, o limitare când este operaționalizat și deci când însușirea este supusă măsurătorilor. Conștientizându-ne acest lucru, nu trebuie să-l considerăm ca un obstacol insurmontabil, ci doar ca un permanent semnal care să ne atragă atenția asupra limitelor cunoașterii ce utilizează asemenea metode, asupra caracterului parțial al concluziilor obținute și deci asupra necesității de a desfășura un permanent travaliu spre desăvîrșirea instrumentelor de analiză.

O scală este fidelă dacă asigură rezultate constante atunci când se aplică în mod repetat aceluiași fenomen. Fidelitatea nu presupune, evident, validitatea căci pot obține mereu cu un instrument aceleași rezultate fără ca el să-mi măsoare ceea ce urmăresc eu. Fidelitatea este deci o condiție minimă, necesară, pe care trebuie s-o îndeplinească orice instrument, fără a fi însă și una suficientă. Testarea fidelității reprezintă o cerință elementară pentru orice scală construită. Nici acest lucru nu este însă chiar atât de simplu în științele sociale, care au de a face cu fenomene schimbătoare și deci nu există certitudinea că la o repetare a scalei se evaluează același lucru ca în momentul inițial.

În încheierea acestor rânduri, țin să precizez, acum după ce, prin succinta prezentare a scalelor, s-au înțeles poate mai bine aspectele abordate în primul paragraf, că insistența mea asupra virtuților analizei cantitative a fenomenelor sociale nu urmărește negarea rolului pe care metodele calitative îl pot juca



în sociologie. Este limpede că nici o știință nu urmărește, în fond, eliminarea calitativului, ci doar a subiectivismului, care în științele umane se ascunde adeseori în spatele unor abordări calitative superficiale și se maschează sau chiar se justifică prin pretinsa imposibilitate a cuantificării fenomenelor sociale. Extrema varietate a acestor fenomene implică și o mare diversitate în posibilitățile de măsurare a lor, deci multiple situații de împletire a cantitativului cu calitativul, demersuri diferite dar care se completează reciproc.

Complexitatea unor fenomene sociale nu trebuie să ne descurajeze în încercarea de analiză riguroasă a lor și să ne arunce în brațele unui impresionism subiectivist. Nu pot desigur măsura, de pildă, satisfacția oamenilor într-o situație oarecare așa cum măsoară intensitatea curentului electric. Pot însă operaționaliza conceptul de satisfacție, pot construi o scală cu două, trei, patru sau mai multe trepte de satisfacție și pot număra câți indivizi ai populației studiate intră în fiecare clasă obținută. Astfel, pot afla ce proporție se află spre partea pozitivă și ce proporție spre cea negativă a scalei, văzînd care situație este predominantă; pot repeta cercetarea și afla în ce direcție evoluează fenomenul; pot compara situația populației mele cu a alteia; pot găsi subclase ale populației în care distribuția pe scală e diferită, ceea ce înseamnă că am detectat un fenomen ce-mi poate explica parțial pe cel studiat etc. Aceștia sînt cîțiva pași modeste, dar fermi în același timp, pe drumul ce duce la cunoașterea obiectivă a realității sociale, la posibilitatea sociologului de a afla ceea ce se ascunde în spatele unor aparențe, de a transcende dincolo de "explicațiile" sociologiei vulgare de care fiecare dintre noi sîntem impregnați.

### 3. Aplicațiile matematicii și statisticii în sociologie

Aplicarea matematicii în științele socio-umane a devenit astăzi o realitate de care nimeni nu se mai poate îndoii. Economia, demografia, lingvistica, psihologia și alte discipline, printre care și sociologia, beneficiază, în măsură mai mare sau mai mică, în funcție de nenumărați factori obiectivi sau subiectivi, de acest



puternic instrument de cunoaștere care e limbajul matematic. Adversarii de principiu ai utilizării matematicii în științele sociale - cel puțin în rândul celor care practică efectiv aceste științe - sînt tot mai puțini; singurele critici cu adevărat obiective ce pot fi aduse respectivei utilizări sînt legate de abuzul de matematizare și cuantificare, de încercările unora de a prezenta, într-o formă prezențioasă și într-un limbaj adesea inaccesibil majorității sociologilor și publicului larg, niște rezultate minore, care ar fi putut fi expuse tot așa de bine și sub o formă mult mai simplă, în limbajul obișnuit.

Mersul înainte al științelor sociale nu se poate realiza, după părerea mea - și, desigur, nu-mi revendic aici nici un dram de originalitate - pe altă cale decît cea parcursă de științele naturii. Or, cum se știe și cum s-a subliniat adesea, acestea din urmă au progresat în măsura în care au fost capabile să facă apel la instrumentul matematic. De la Bacon încolo, numeroși oameni de știință și filosofi au subliniat corelația ce există - sau trebuie să existe - între spiritul științific și cel matematic. Sociologia, ca nici o altă știință, în măsura în care se vrea știință, iar termenul acesta nu e golit de orice conținut, nu poate face excepție de la o regulă ce s-a manifestat atît de puternic pînă acum.

Analizînd situația de fapt, existentă azi în sociologie, putem constata ușor că aparatul matematic este aici mai puțin prezent, iar rezultatele obținute pe această cale mai puțin spectaculoase, decît în alte științe, chiar sociale. Încercări de a explica această situație sînt destule, mergînd de la reproșul făcut matematicienilor că nu dezvoltă suficient acele capitole din domeniul lor care se pretează la aplicații în sociologie și pînă la cele adresate sociologilor că sînt insensibili la spiritul matematic, rămînînd la o formație umanistă desuetă sau pur și simplu că nu sînt capabili să priceapă nimic din matematică. În cadrul acestei discuții revine mereu, ca un leitmotiv, problema enunțată în paragrafele precedente, și anume cea a raportului dintre cantitativ și calitativ. Se argumentează că fenomenele sociale se pretează greu la cuantificări și deci acesta ar fi motivul pentru care matematicile, prin excelență științe ale cantității, nu-și găsesc cîmp



suficient de larg de aplicare în sociologie. Se argumentează și invers, că matematica a încetat de a folosi doar elemente cantitative, că laturile sale calitative ar fi cele preponderente azi și deci că sociologii n-au decît să caute și vor găsi ceea ce le este necesar.

O discuție și o polemică la acest nivel atît de abstract mi se pare cu totul inutile. După ce am precizat cît de cît înțelesul termenilor de cantitativ și calitativ, pentru a sugera unele răspunsuri la întrebări pertinente, se cuvin cîteva cuvinte asupra înțelesului expresiei aplicarea matematicii în sociologie. Voi observa, mai întîi, că uneori se folosește ca termen echivalent cel de matematizare a sociologie, dar cu o frecvență totuși mai mică, poate și pentru că este puțin mai tare, sugerînd o infiltrare mai puternică a matematicii în sociologie, golirea acesteia de conținutul său concret.

S-a subliniat adesea distincția de esență ce există între științele logico-matematice și științele nomotetice sau ale realului sau empirice, indiferent cum le vom numi. Dacă această distincție este corectă - și cred cu toată convingerea că lucrurile stau așa -, atunci rezultă, printre altele, că matematica nu poate fi altceva pentru sociologie decît o metodă, un instrument, un auxiliar, un limbaj. Cu alte cuvinte, nu e nici un pericol de dispută între aceste științe pentru vreo chestiune teritorială. Asemenea diferende pot apare doar între științele realului care se învecinează pe porțiuni comune; matematica nu se învecinează cu nici una, fiind, în același timp, în relație cu toate.

Dacă acceptăm concepția curentă conform căreia orice știință nomotetică este structurată pe două niveluri principale - cel empiric și cel teoretic -, atunci putem constata că matematica își găsește posibilități de aplicare pe ambele paliere. La nivelul cercetării sociologice empirice, matematica permite o analiză a datelor sociologice culese, o prelucrare a lor, și mai ales, construirea unor modele și scheme explicative.

Pentru exemplificare, să presupunem că între mai multe variabile am stabilit o relație funcțională de genul:

$$y = f(x_1, x_2, \dots, x_n)$$



studiere matematică poate contribui la realizarea câtorva cognitivi:

a) Evidențierea raporturilor cantitative existente între variabila dependentă,  $y$ , și cele independente.

b) Detectarea influenței specifice a fiecărui factor  $x$  asupra lui  $y$ . Astfel, stabilind că pe un domeniu oarecare funcția este derivabilă în raport cu  $x_1$  și că, de exemplu,

$$\frac{\partial f}{\partial x_1} > 0$$

atunci vom putea spune că, în limitele domeniului unde se realizează inegalitatea de mai sus, o creștere a lui  $x_1$  antrenează o creștere a lui  $y$ , ceilalți factori rămânând constanți. Si în mod similar se poate proceda cu toate variabilele independente.

c) Formularea unor previziuni în legătură cu  $y$ . Cunoșcând tendința de variație a fiecărui  $x$ , prin combinarea acestora în funcția  $f$ , va rezulta tendința lui  $y$ .

La nivel teoretic, aplicațiile matematicii în cadrul unei științe înseamnă, în esență formalizarea limbajului său. Cât privește sociologia, consider că deocamdată această problemă nu se pune, cel puțin la nivelul teoriei generale. În cadrul anumitor teorii parțiale sau de rang mediu, limitate la o sferă de fenomene simple, s-au realizat încercări de formalizare, dar ele nu au reușit încă să-și demonstreze pe deplin viabilitatea.

Se poate, cred, aprecia că, în momentul de față, încercările de aplicare a matematicii în sociologie au un efect benefic pentru ambele discipline. Sociologia, chiar dacă nu realizează profeții explicative deosebite, va fi pozitiv influențată în dezvoltarea sa cel puțin pe linia precizării conceptelor. Se cunoaște rigoarea limbajului matematic și, în același timp, imprecizia ce domină în cel sociologic. Întâlnirea celor două discursuri nu se poate face atîta vreme cît nu se aruncă o punte peste această adevărată prăpastie. Or, puntea trebuie să fie aruncată de pe tărîmul sociologiei căci matematica nu poate renunța, fără a se nega pe ea însăși, la exactitatea conceptelor sale.

Pe de altă parte, matematica poate beneficia de aplicații în domeniul social, în scopul extinderii și diversificării teoriilor



sale. Este adevărat că dezvoltarea de pînă acum a domeniilor matematice a avut drept stimulent extrinsec aplicațiile în domeniul științelor naturii sau tehnice. De aceea e plauzibil că, cel puțin parțial, ea și-a concentrat activitatea spre a răspunde cît mai bine acestor cerințe, influențîndu-se astfel structura, raportul de forțe între diferitele ramuri ale sale. Nu trebuie totuși exagerat acest aspect, nu trebuie supraestimată o atare influență. Multe descoperiri matematice n-au avut nici o legătură practică imediată, iar, dacă și-au găsit-o, asta s-a întîmplat peste decenii sau secole. Domeniul matematic este relativ autonom, avînd o logică a sa internă de dezvoltare.

Legat de problema de a spune ce tip de matematici, cantitative sau calitative, se pretează cel mai bine sociologiei, nu se pot face decît afirmații ipotetice. Este o chestiune pe care viitorul o va decide. Un lucru este însă sigur, și anume că, în ciuda dezvoltării matematicilor nonmetrice, dacă se face o analiză a ceea ce s-a aplicat pînă în prezent în sociologie, rezultă că ramurile cantitative au dus la rezultatele cele mai importante și au avut, de departe, cele mai frecvente utilizări. Lucrul este explicabil și prin legătura strînsă ce există între matematică și statistică, celălalt instrument fundamental utilizabil pentru ridicarea valorii rezultatelor cercetărilor sociale.

Statistica este o disciplină a cărei naștere, chiar dacă nu poate fi exact datată, se plasează cu siguranță în secolul XVII. Ea s-a dezvoltat inițial pe două direcții, în virtutea celor două puncte de plecare, diferite în spațiu și prin problematică, dar aproape simultane în timp. Scoala germană a dezvoltat ceea ce s-ar putea numi "statistica statală", o disciplină chemată să contabilizeze, ordoneze și să analizeze date prin care poate fi caracterizată dezvoltarea statului într-o etapă sau alta. Scoala engleză - cu reprezentanți mai cunoscuți astăzi precum: John Graunt, William Petty sau celebrul astronom Halley - a dezvoltat ceea ce s-a numit inițial "aritmetica politică", de fapt demografia bazată pe instrumentul statistic. Aparent, școala germană, prin obiectivele sale mai largi, ce ofereau o paletă extinsă de aplicații pentru noua tehnică, ar fi trebuit să constituie cadrul cel mai favorabil de dezvoltare a statisticii. Lucrurile nu s-au întîmplat așa deoarece discursurile despre stat și dezvoltarea lui erau în mod



precumpănitor verbale, aspectul cantitativ-cifric nereușind să înlocuiască aprecierile impresionist-speculative. În schimb, școala engleză, limitată la o arie mai restrânsă de fenomene, a dezvoltat metodologia statistică în punctele sale și azi esențiale, fiind, de cele mai multe ori, recunoscută drept singur precursor.

Oricum, nu este în intenția mea de a face aici istoria statisticii; unicul lucru pe care doresc să-l subliniez cu toată tăria pentru a fi reținut este faptul că statistica s-a născut și s-a dezvoltat pe problematica științelor sociale. Simbioza aceasta a fost atât de puternică încât până la sfârșit secolului trecut, când s-au amplificat utilizările ei și în alte domenii, statistica a fost considerată drept o disciplină socială de sine stătătoare. Deci foarte târziu s-a remarcat caracterul metodologic al disciplinei, foarte târziu s-a văzut că, de fapt, ea nu are un obiect propriu de studiu și că metodele sale pot fi aplicate în orice disciplină științifică, indiferent de domeniul - natural ori social - al obiectului său. Astfel s-a constituit un nucleu identic unitar ce poate fi numit "statistica generală" sau "teoria statisticii", format din totalitatea conceptelor, procedurilor și operațiilor statistice general valabile. Pe lângă acestea există desigur concepte și proceduri specifice aplicațiilor în anumite științe particulare, cum sînt economia, demografia etc.

Fără pretenția de a da o definiție, se poate spune că statistica este o disciplină ce se distinge prin cel puțin trei elemente esențiale:

a) caracterul metodologic amintit deja, adică lipsa unui obiect propriu și faptul că ea constă dintr-un arsenal de tehnici aplicabile în alte discipline;

b) fenomenele, însușirile etc. studiate cu mijloace statistice se relevă ca aspecte variabile pe o masă mare de unități (indivizi);

c) aceste fenomene sînt afectate de o pluralitate de cauze în manifestarea lor și, ca atare, relațiile esențiale se evidențiază ca o rezultată a "comportamentelor" individuale.

Aplicațiile statisticii în domeniul sociologiei sînt multiple mergînd de la oferirea unor modalități de analiză a unei însușiri cu manifestare cantitativă sau calitativă (evidențierea



distribuției indivizilor în diferite clase, calculul procentelor, mediei, abaterii standard etc.) și pînă la construcția unor modele sofisticate de detectare a relațiilor dintre mai multe variabile, oferind astfel argumente obiective pentru încercările de a explica fenomenele sociale. Acest lucru a fost sesizat de la bun început, nașterea statisticii producîndu-se, cum am amintit, tocmai pe tărîmul faptelor sociale. Cele trei secole ce au trecut de la debutul noii perspective de abordare a fenomenelor sociale n-au făcut decît să întărească convingerea în validitatea metodelor statistice, metode care n-au încetat să se amelioreze, să se rafineze, mai ales de la sfîrșitul secolului trecut încoace. Mijloacele electronice de stocare și prelucrare a informației, apărute în scenă la mijlocul secolului nostru au dat un imbold extraordinar aplicațiilor statisticii în științele sociale, iar sociologia a fost unul din beneficiarii care au profitat cel mai mult de noile condiții.

Relațiile statisticii cu matematica sînt foarte strînse, nu numai pentru motivul, ușor de detectat și de înțeles, că ambele lucrează cu cifre. O serie de componente esențiale ale statisticii și-au găsit o descriere și o fundamentare matematică - în speță în domeniul teoriei probabilităților. Intensitatea legăturilor dintre cele două discipline se observă foarte clar în aplicații, cînd, deseori, este greu de spus unde se termină statistica și intră în lucru matematica. Cu toate acestea, ar fi greșit să se considere statistica drept o ramură a matematicii. Distincția dintre cele două discipline devine evidentă la nivelul discursului în teoria statisticii, unde lipsește cu devărvîrșite caracterul axiomatic, deductiv chiar, trăsături specifice teoriilor matematice. Apoi, în ciuda faptului că unele noțiuni statistice au un corespondent matematic, ele nu coincid cu aceste imagini ideale ale lor (noțiunea de frecvență relativă nu este identică cu cea de probabilitate, curbele de frecvențe empirice nu reproduc niciodată pe cele ideale etc.). Mai mult, o serie de noțiuni statistice - indicatori, indici, coeficienți etc. nu au nici o justificare matematică precisă, ei fiind rezultatul exclusiv al demersului statistic, al analizei datelor empirice. De aceea nu este de mirare că pentru a măsura același lucru întîlnim, în statistică, trei, patru sau chiar mai multe mărimi propuse; o asemenea "dezordine" nu este



evident, posibilă în matematică. O imagine asupra unor instrumente de analiză statistică va fi oferită în ultimele două părți ale cursului de față. În încheierea acestui capitol, voi preciza doar câteva noțiuni de bază, ele fiind, de fapt, rezultatul întregii discuții de până aici. Așa cum am spus, statistica lucrează cu mase de indivizi, plecând de la descrierea modului în care variază diferitele fenomene în rândul acestora. Aceasta înseamnă - după terminologia folosită în primul paragraf - că statistica, deși are ca obiect de analiză însușirea, utilizează exclusiv perspectiva cantitativistă la nivelul obiectului-suport.

Drept consecință, noțiunile primare ale statisticii sînt cele de indivîd sau, echivalent, unitatea statistică, reprezentînd obiectele-suport purtătoare de însușiri și de univers (sau echivalent: populație, colectivitate statistică), în calitate de mulțime a tuturor obiectelor pe care se studiază o însușire dată. Termenul de individ - și, corelat, cel de populație - nu trebuie să ne inducă în eroare. Nu e vorba de (sau numai de) indivizi umani; această terminologie va fi aplicată oricărei unități de analiză: familii, grupuri, orașe, sate, instituții, națiuni etc. Individul statistic este deci ceea ce se numără, iar însușirea - sau caracteristica, în limbaj statistic consacrat - este cea care se evaluează cu o anumită scală. Se vor distinge, prin, urmare, metode statistice calitative sau nonparametrice, dacă scala este nominală sau ordinală, și metode statistice cantitative, cînd ea e de intervale sau rapoarte, deci cînd se realizează măsurarea însușirii. Așadar, atributul de cantitativ aplicabil unor proceduri statistice vizează în exclusivitate trăsătura, deoarece la nivel de obiect abordările statistice sînt prin definiție cantitative.

Ca și în cazul matematicii, folosirea mijloacelor de analiză statistică în sociologie - făcută cu discernămint și fără exces - nu poate avea decît efecte pozitive asupra limbajului, raționamentului, și în general, asupra întregului demers descriptiv-explicativ al acestei științe.



METODE DE CULEGERE A INFORMATIEI

Cap. IV. ANCHETA

Fără teama de a greși, se poate aprecia că ancheta este metoda de cercetare cu cea mai largă utilizare în sociologia empirică. Această frecvență ridicată de apariție a metodei în studiile "de teren" n-a întîrziat să creeze o imagine falsă în ochii nespecialiștilor, care o concept ca metodă exclusivă a sociologiei, iar sociologul este adesea reprezentat, într-o imagine populară, drept "omul cu chestionarul". Indiferent de criticile destul de severe adresate de-a lungul timpului acestei modalități de culegere a informației, ancheta a rămas și astăzi o metodă de bază, iar literatura metodologică referitoare la ea a proliferat atît de mult încît orice încercare, ca cea de față, de a concentra în cîteva pagini ideile principale legate de utilizarea sa reprezintă, pur și simplu, un act de temeritate și, desigur, implică riscul de a omite aspecte deosebit de importante.

Ancheta, ca metodă specifică sociologiei, a spărut din particularitatea pe care unele dintre "obiectele materiale" de cercetare (oamenii) o au, și anume aceea de a permite un schimb direct de informație între cercetător și "obiect", prin intermediul limbajului. Renunțarea la ancheta ar însemna, în ultimă instanță, renunțarea la acest avantaj al comunicării interumane, deci la privarea sociologiei de o sursă de informații, care, chiar dacă ridică, de cele mai multe ori, probleme metodologice-tehnice și interpretative foarte dificile, îi poate aduce un inedit beneficiu cognitiv.

Perfecționările metodice realizate pînă în prezent - și posibil de continuat în viitor - au condus la creșterea gradului de validitate a cunoștințelor dobîndite prin intermediul anchetei, astfel încît deja se poate vorbi realmente de ea ca de o modalitate fundamentală în studierea, faptelor, fenomenelor și chiar a proceselor sociale. Din păcate, publicul larg cunoaște foarte puține



din dificultățile ridicate de utilizarea efectivă a metodei și nu rareori ea este discreditată de încercări ale amatorilor, care văd aici o posibilitate extrem de facilă de a întreprinde ceva capabil să le eticheteze activitatea cu atributul, atât de râvnit, de "științifică". De aceea, în capitolul de față, voi încerca - din cauza întinderii neobișnuite a problematicii - să abordez doar unele aspecte, capabile să creeze, sper, o imagine de ansamblu a chestiunilor care se ridică, să atragă atenția asupra dificultăților metodico-practice și, evident, să convingă pe cel care ar intenționa să activeze pe tărîmul cercetării sociale să se îndrepte mai întâi spre lectura unor lucrări.

### 1. Metoda și tehnicile sale

În primul capitol al cursului am prezentat o clasificare a metodelor sociologiei, în rîndul căroră am inclus ancheta și interviul, ca metode distincte, sugerînd în treacăt și deosebirea dintre ele. Este locul să reluăm problema, mai cu seamă că forma de clasificare pe care am utilizat-o impune de la sine o definire a fiecărei metode numai prin raportare la specificul celeilalte.

Fără pretenția unor definiții riguroase, vom admite că ancheta este metoda de culegere a informației de la o masă mare de indivizi umani, cu ajutorul unui instrument specific: chestionarul. La rîndul său, chestionarul poate fi înțeles ca o mulțime de întrebări, de regulă scrise, adresate subiecților în forme variate, într-o anumită succesiune, pe baza unor considerente logice și psihologice.

Din punct de vedere tehnic, de realizare practică efectivă, metoda anchetei are două variante fundamentale;

- a) ancheta directă sau orală și
- b) ancheta indirectă sau în scris.

În primul caz, întrebările chestionarului se adresează subiecților anchetați sub formă orală, de către cercetător sau de persoane auxiliare folosite de acesta doar pentru culegerea informației, numite operatori de anchetă. În al doilea, sociologul ia doar un contact indirect cu subiectul, întrucît acesta din urmă este chemat să citească și să completeze singur chestionarul-formular care i se



înmînează într-o formă sau alta.

Tehnica anchetei directe se realizează, de cele mai multe ori, printr-un dialog față în față între cercetător (operator) și subiectul supus investigației (sau, cum se spune în franceză, prin "entretien personnel"), dar, în anumite cazuri, cu totul deosebite, ea poate lua și forma unui dialog prin telefon. Dat fiind faptul că ancheta prin telefon este foarte rar folosită, nu voi mai reveni în continuare asupra ei, menționînd aici că ea se realizează doar pentru colectarea unei cantități reduse de informație și de la un public nu prea numeros. Mai mult, aria problematicii astfel abordate este restrînsă, deoarece nici în cele mai dezvoltate țări telefonul nu se găsește în fiecare casă; deci populația vizată de cercetare trebuie să fie inclusă în populația care posedă acest mijloc de comunicare.

Ancheta indirectă se poate efectua în mai multe forme de realizare concretă. Una dintre cele mai vechi, deci dintre primele utilizate, constă în publicarea chestionarului în ziare și reviste și colectarea răspunsurilor tuturor celor care doresc să returneze formularul completat. Astăzi acest procedeu se mai practică doar în scopuri publiciste și publicitare, cercetările științifice serioase renunțînd de mult la el, pentru motivul principal că, oricît de mare ar fi numărul chestionarelor returnate - și el ajunge adesea la ordinul zecilor de mii -, persoanele care răspund la asemenea solicitări nu se constituie într-un eșantion cu o reprezentativitate suficient de ridicată spre a putea generaliza, cît de cît corect, constatările asupra întregii populații.

O altă formă de materializare a tehnicii anchetei indirecte constă în strîngerea unei mulțimi de persoane într-o sală în scopul completării chestionarului. De regulă, operația se realizează sub controlul cercetătorului, care distribuie chestionarele, explică scopul anchetei și precizează modalitatea de completare. Intervenția cercetătorului poate lua forme variate, mergînd de la cîteva indicații tehnice prealabile și supravegherea subiecților pentru a nu "colabora" la completare pînă la citirea fiecărei întrebări în parte și chiar explicarea sau comentarea ei. Pe această cale se poate deci ajunge atît de departe încît să se iasă efectiv din cadrul tehnicii de anchetă indirectă propriu-zisă, fie în direcția unei anchete orale sui-generis, fie spre o tehnică psihoso-



ciologică consacrată sub numele de interviu de grup. Bineînțeles că forma de anchetă colectivă se poate realiza doar în cazul când sînt îndeplinite unele condiții speciale. Pe de o parte, este nevoie ca procedura de eșantionare să fie adecvată acestui scop (eșantionare multistadială, de regulă), iar, pe de altă, să avem efectiv posibilitatea să reunim mai multe persoane într-o sală, ceea ce este mai simplu atunci când anchetăm oamenii la locul de muncă, în școli, facultăți etc.

A treia specie a anchetei indirecte se realizează prin înmînarea chestionarului, în mod individual, fiecărui subiect de către un membru al echipei de cercetare sau auxiliar. Și aici întâlnim un avantaj larg de variante, ce corespund în mare celor din cazul precedent. Adică, la un pol se poate imagina situația în care cercetătorul realizează doar un contact superficial cu subiectul atît cît să-i fie necesar să-l determine la completarea chestionarului și să-l lămurească asupra aspectelor tehnice ale înregistrării răspunsurilor, iar la altul cînd ancheta indirectă abia mai poate fi distinsă de cea directă, într-atît de implicat devine cercetătorul în operația de completare a formularelor. În situațiile apropiate de forma mai impersonală, cele mai des întîlnite, chestionarul - după lămuririle prealabile - poate fi lăsat în posesia persoanei chestionate mai multe zile, urmînd a se reveni pentru ridicarea lui.

În sfîrșit, o a patra modalitate de anchetă indirectă este cea realizată prin poștă: chestionarele se trimit subiecților ce formează eșantionul, însoțite de o scrisoare în care se precizează scopul anchetei, felul cum s-au ales persoanele anchetate și se dau instrucțiunile de completare. De asemenea, prin cele mai politicoase și persuasive formule, se va solicita returnarea chestionarului completat, în care scop se va atașa un plic timbrat cu adresa instituției ce efectuează cercetarea.

Din sumara trecere în revistă a diverselor tehnici de anchetă, cred că se poate desprinde specificul metodei, fiind mai ușor acum de precizat diferența față de interviu. În esență, ancheta, indiferent de modalitatea concretă de realizare, presupune, așa cum am precizat, apelul la un chestionar, adică la un set de întrebări relativ clar structurate, ce se adresează, direct sau indirect, subiecților, într-o formă standard, sau în orice caz, apropiată



tă de aceasta. Interviuul, chiar dacă face apel la un instrument similar chestionarului, numit ghid de interviu, se deosebește de anchetă prin aceea că cercetătorul realizează în fapt o discuție relativ liberă cu subiectul, ghidul de interviu având drept unic scop să-l ajute pe cel care conduce convorbirea să o mențină în cadrele inițial fixate. Se vede deci că interviuul poate fi apropiat de ancheta directă și nu puține manuale de metode și tehnici le tratează indistinct. Mai exact, se vorbește de o aceeași metodă care ar lua forme diverse, mergând pe o anumită scală de formalizare de la discuția absolut liberă până la cea perfect structurată, pe baza unui chestionar care nu permite nici o înregistrare necodificată a informației furnizate de subiect și nici o libertate pentru anchetator în a schimba textul întrebărilor și ordinea lor.

Este adevărat că cele două metode interferează, că în unele situații ancheta poate lua turnură de interviu, dar cred că mai rațională este distingerea lor așa cum am făcut-o în paginile de față. Pe lângă instrument, un alt factor care pledează pentru o atare distincție este și numărul subiecților investigați. Într-adevăr, în cazul interviurilor acest număr nu poate fi decât mic sau foarte mic, deoarece, pe de o parte, durata unui interviu este net superioară celei a unei anchete, apoi, pe de altă parte, numărul persoanelor capabile să realizeze interviuri este inferior celui al operatorilor ce pot fi antrenați în anchete, deoarece pentru aceasta din urmă formă nu e nevoie de o calificare deosebită (practic interviul nu poate fi făcut cu personal auxiliar) și, în fine, prelucrarea datelor unui număr mare de interviuri este, dacă nu imposibilă, extrem de greu de realizat, având în vedere caracterul absolut nestandardizat al formelor de înregistrare: text liber, scris ori stenografiat, sau bandă de magnetofon. Ancheta, dimpotrivă, tocmai datorită chestionarului, poate fi efectuată pe un eșantion mare de persoane, căci implică riscuri mult mai reduse în folosirea anchetatorilor nespecialiști și oferă posibilitatea unor prelucrări rapide ale datelor, mai ales acum în era calculatoarelor electronice.

O consecință firească a modului de realizare constă în faptul că interviul permite abordarea problemelor în profunzime,



discuția putînd fi purtată asupra celor mai mici detalii sau a celor mai subtile aspecte psihologice. Ancheta este, prin natura lucrurilor, condamnată să se mărginească la aspecte mai simple, mai ușor de investigat, la probleme ce pot fi abordate într-o discuție rapidă sau pe care oamenii sînt dispuși să le clarifice în cadrul unui formular ce le este pus în față spre completare. De aici rezultă deci avantajele și dezavantajele fiecărei metode: ancheta este superioară datorită numărului mare de indivizi chestionați, ceea ce asigură o reprezentativitate ridicată eșantionului, deci garanția că datele obținute sînt relevante pentru întreaga populație; interviul vine cu argumentul calității și al profunzimii, cu o mai largă deschidere spre tema studiată. În aceste condiții mi se pare natural ca, așa cum am mai spus-o, cele două tehnici să se completeze reciproc. Forma optimă de împletire a lor, recomandată în unele lucrări de specialitate, ar fi următoarea. Cercetarea ar trebui să înceapă cu un interviu al unui lot restrîns de subiecți, ceea ce va permite cercetătorului familiarizarea cu problema studiată, descoperirea unor aspecte, dimensiuni nebanuite ale fenomenelor abordate, schițarea unor posibile scheme explicative sau îmbogățirea arsenalului de ipoteze etc. A doua fază ar constitui-o ancheta pe un lot mare și cu instrumente standardizate, a căror construcție este evident facilitată de interviurile de etapă precedentă. În fine, pornind de la concluziile degajate din anchetă, se recomandă reluarea interviurilor (pe același lot ca și la început sau pe altul), de data aceasta cu scopuri mult mai clar conturate, pentru a întregi constatările generale, pentru a găsi explicații adecvate unor corelații formale etc.

După această scurtă paranteză, să revenim la tehnicile de anchetă, să vedem ce probleme practice ridică fiecare, ce avantaje sau dezavantaje însoțesc alegerea unei anumite tehnici într-o cercetare sociologică efectivă.

Mai întîi o remarcă: manualele de metode și tehnici par să încline în a acorda un avantaj decisiv în formarea anchetelor directe (orale), avantaj ce ar rezulta, așa cum foarte concis se exprimă Caplow, din faptul că oamenii găsesc mult mai multă plăcere în a purta o conversație decît în a completa un formular (1970, p.222). Ideea aceasta, în forme mai mult sau mai puțin sofis-



ticate, se regăsește în bună parte din lucrările celor care au scris despre anchete. De fapt, se poate întâlni uneori și o notă de surpriză legată de succesul pe care-l au operatorii de anchetă pe lângă persoanele chestionate. Dacă cei dinții sînt bine pregătiți, au o atitudine și un comportament adecvate față de subiecți, ei pot ajunge la a aborda cele mai delicate probleme din viața intimă a indivizilor, așa cum au demonstrat-o anchetele legate de comportamentul sexual (Kinsey).

Madeleine Grawitz (1972, p.653) consideră că, în general, succesul anchetei ar putea fi explicat pe baza a trei factori principali, dintre care ultimul joacă efectiv în favoarea celei directe:

a) Reflexul de politețe. Este greu să refuzi să-i vorbești cuiva care ți se adresează politicos, te roagă să-l ajuți în realizarea muncii ce-o are de îndeplinit etc.

b) Dorința de influențare. Cu cît subiectului i se sădește o convingere mai puternică despre faptul că prin răspunsurile sale va putea contribui la schimbarea unor situații, la ameliorarea unor stări de lucruri care-l privesc direct, el va fi dispus într-o măsură mai mare să se angajeze în discuție.

c) Nevoia de a vorbi. E vorba nu de sensul vulgar de pălăvrăgeală, ci de unul mai adînc, psihologic, de nevoia de comunicare și de a fi înțeles. Dacă n-ar exista aceste nevoi, oricîtă politețe și bunăvoință ar manifesta subiectul, ar fi totuși de neconceput ca o persoană singură acasă (mai ales în mediul urban) să deschidă ușa unui necunoscut spre a-i răspunde la întrebări, trecînd peste toate riscurilor pe care un astfel de gest le poate implica.

Așadar, tehnica anchetei directe se dovedește a fi mai atractivă pentru subiecți, de unde rezultă unul din avantajele sale: ponderea nonrăspunsurilor este, în general, mai redusă decît în ancheta indirectă. Intr-adevăr, aceasta din urmă, mai ales dacă este vorba de chestionarul prin poștă, ridică nenumărate obstacole în calea unei interpretări corecte a rezultatelor tocmai din cauza numărului mare de persoane care nu returnează chestionarul. În felul acesta, rezultatele anchetei sînt, în principiu, distorsionate dat fiind că, foarte probabil, persoanele care nu răs-



pund au alte caracteristici decât cele care răspund. În consecință nu va fi suficientă doar o supradimensionare a eșantionului inițial în speranța că prin nonrăspunsuri se va obține eșantionul de mărime dorită; oricât de mulți indivizi am include, gradul de distorsiune nu se reduce, căci el depinde de proporția nonrăspunsurilor și nu de numărul acestora.

Ideea de mai sus poate primi o formulare matematică foarte clară. Să notăm cu  $p$  proporția persoanelor din eșantion care răspund la chestionar și cu  $1-p$  a celor care nu răspund. Să mai notăm cu  $\bar{x}$  proporția indivizilor din eșantionul total care posedă o anumită caracteristică, cu  $\bar{x}_1$  proporția corespunzătoare în rândul celor care răspund și cu  $\bar{x}_2$  în subeșantionul celor care refuză să răspundă. Atunci proporția  $\bar{x}$  se poate scrie:

$$\bar{x} = p\bar{x}_1 + (1-p)\bar{x}_2$$

Efectuând cercetarea și obținând numai o parte din chestionare, vom ajunge să determinăm pe  $\bar{x}_1$  în loc de  $\bar{x}$ , sau, altfel spus, se va produce o distorsiune  $d$  egală cu diferența dintre cele două valori:

$$d = \bar{x} - \bar{x}_1$$

Înlocuind pe  $\bar{x}$  din formula precedentă, obținem:

$$d = p\bar{x}_1 + (1-p)\bar{x}_2 - \bar{x}_1 = (1-p)\bar{x}_2 - (1-p)\bar{x}_1$$

sau

$$d = (1-p)(\bar{x}_2 - \bar{x}_1)$$

Prin urmare, distorsiunea introdusă este direct proporțională cu  $1-p$ , adică cu ponderea nonrăspunsurilor, și cu  $\bar{x}_2 - \bar{x}_1$ , deci cu diferența de comportament între indivizii care răspund și cei care nu răspund. Distorsiunea se va anula atunci când unul dintre factori este zero. Or, cum, de regulă, despre al doilea factor noi nu avem suficiente cunoștințe, nu vom putea presupune decât cu multe riscuri că el este nul. Rămîne, așadar, ca mult mai sigură calea de anulare a primului factor, adică străduința de a elimina pe cît posibil nonrăspunsul.



Calculul schițat mai sus este, evident, valabil pentru orice gen de anchetă, ceea ce susține afirmația că problema non-răspunsului este una generală. Numai că gravitatea ei diferă în funcție de diversele forme pe care ancheta le poate lua. Altfel spus, nonrăspunsul apare în toate anchetele, dar proporțiile în care el se prezintă variază foarte mult de la un tip de anchetă la altul, minimul fiind înregistrat în cea directă iar maximul în cea indirectă prin poștă.

Un anumit procent de nonrăspunsuri nu poate deci fi eliminat în nici o anchetă, dar dacă valoarea sa este mică, atunci distorsiunea produsă este și ea scăzută. Să exemplificăm puțin lucrurile. Să presupunem că trăsătura urmărită apare la 70% dintre indivizii care răspund și doar la 50% din cei care nu răspund, adică cu notațiile de mai sus:

$$x_1 = 0,70$$

$$x_2 = 0,50$$

Dacă proporția nonrăspunsurilor este de 10% ( $1-p=0,10$ ), cifră ce caracterizează unele anchete directe, atunci distorsiunea va fi:

$$d = 0,10 \cdot 0,20 = 0,02$$

decî o valoare de numai 2 procente, ce se încadrează, de regulă, în marja de eroare acceptabilă în astfel de cercetări. Dimpotrivă, dacă cei care nu răspund sînt în proporție de 80% ( $1-p=0,80$ ), așa cum se întîmplă la chestionarul prin poștă, distorsiunea va primi valoarea:

$$d = 0,80 \cdot 0,20 = 0,16$$

adică 16 procente, ceea ce va afecta serios concluziile obținute.

Un al doilea avantaj al anchetei directe îl constituie faptul că, de cele mai multe ori, în asemenea situații persoanele interogate acordă o atenție mai mare chestionarului. Faptul că răspunsul este dat unei persoane aflată în fața ta are desigur o anumită influență asupra griiei de a înțelege corect fiecare întrebare, de a nu te contrazice, și în general, de a privi cu seriozitate întreprinderea respectivă. Avînd doar chestionarul în față



ești tentat să-l completezi cît mai repede pentru a te scăpa de o grijă, de unde o anumită neglijență, superficialitate etc. Desigur, lucrurile nu stau întotdeauna astfel, căci există anumite categorii de persoane care sînt mai conștiincioase în completare decît operatorii înșiși, dar ele nu reprezintă decît excepțiile care confirmă regula.

Un al treilea avantaj ce poate fi invocat pentru ancheta directă este spontaneitatea cu care se dau răspunsurile la chestionar în acest caz. Problema spontaneității este relevantă în cazul întrebărilor la care se cer opiniile, părerile, atitudinile oamenilor și nu atît anumite fapte. Intr-adevăr, pentru primul gen de întrebări este deosebit de important ca răspunsurile să fie provocate cît mai direct posibil. O perioadă de meditație, facilitată de posesia chestionarului, conduce, de regulă, la alterarea opiniilor respective, fie în sensul că omul are tendința de a se consulta cu semenii săi apropiați (și atunci se amestecă opiniile mai multora), fie că va medita asupra eventualelor consecințe ce le-ar putea avea pentru el emiterea unor păreri și va tinde să se conformeze unor opinii dezirabile din punct de vedere social. De altfel, chiar și fără răgaz de gîndire opinii false (spre deosebire de fapte false) sînt mai ușor de exprimat în scris decît în fața cuiva, cel puțin din motive de jenă.

În fine, anchetele indirecte sînt posibile numai în cazurile în care nu este nici un pericol de contaminare a răspunsurilor la diferite întrebări, contaminare ce se produce atunci cînd subiectul are posibilitatea să citească dinainte toate întrebările chestionarului. De regulă, în chestionar - mai cu seamă la întrebările de opinie - are importanță și ordinea în care acestea sînt adresate, recurgîndu-se la o serie de stratageme pentru a obține și pe această cale răspunsuri cît mai corecte. Or, în ancheta indirectă astfel de artificii sînt imposibile.

Cu siguranță că și pentru anchetele indirecte se pot pune în evidență o serie de avantaje, altfel ele ar fi dispărut din practica cercetării sau nici n-ar fi fost inventate. În primul și în primul rînd aceste valanete sînt de ordin material. Costul cercetării în anchetele indirecte - luat termenul de cost într-un sens larg ce implică nu numai mijloace bănești dar și umane - este mult redus, căci fie se elimină complet deplasarea operatorului la su-





biect (chestionarul prin poştă), fie un cercetător anchetează simultan un număr mare de subiecţi (anchete colective cu autocompletarea chestionarului), fie timpul pierdut chiar cu aplicarea individuală este redus la minimum.

În al doilea rând, nu este imposibil să se imagineze şi situaţii în care ancheta indirectă aduce şi avantaje de ordinul cunoaşterii. Într-adevăr, dacă informaţia cerută este de natură factuală, ea poate presupune un anumit timp de gândire sau chiar o activitate prealabilă de documentare din partea subiectului, lucruri ce pot fi mai lesne făcute dacă operatorul nu este de faţă. Dar trebuie s-o recunoaştem astfel de situaţii sînt mai rare şi, din punctul de vedere al acurateţii datelor, este preferabil de cele mai multe ori ancheta directă. Afirmatia devine şi mai convingătoare dacă ne gândim că, în ciuda eliminării analfabetismului, un procent nonneglijabil din populaţie nu are deprinderi deosebite de scris şi de citit. A pune în faţa unor asemenea oameni un chestionar cu 2-300 întrebări nu înseamnă numai a-i expune unei activităţi dificile, ci şi a te expune pe tine ca cercetător riscului ca la o parte din întrebări să nu fie înregistrate corect răspunsurile, ca o altă parte să nu fie corect înţelese sau ca, din cauza efortului, de la un moment dat subiectul să treacă pur şi simplu prin chestionar fără nici un interes sau chiar să abandoneze completarea lui.

Din toate cele de mai sus, cred că rezultă cu claritate că ancheta indirectă poate fi utilizată atunci cînd avantajele materiale impun acest lucru, dar numai cu condiţia ca eventualele pierderi de ordin cognitiv să fie minime. În general, chiar dacă tipul de informaţie solicitat şi chestionarul întocmit nu ridică probleme în ceea ce priveşte autocompletarea, este recomandabil să se recurgă la această tehnică numai în cazul în care populaţia vizată de cercetare posedă în întregime un anumit nivel de instrucţie (e preferabil să aibă cel puţin liceul terminat). În caz contrar, riscurile de erori şi nonrăspunsuri sînt atît de mari încît economiile realizate nu pot justifica adoptarea tehnicii anchetei indirecte.

Datorită faptului că în anchetele directe un rol important revine operatorilor, nu numai ca elemente intermediare, mijlocitoare în procesul de circulaţie a informaţiei de la subiectul investigat la cercetător, deci nu numai ca un, să zicem, conductor



pasiv, ci, în primul rînd, ca elemente active care provoacă emi-  
terea informației, mă voi opri, în cîteva cuvinte, asupra activi-  
tății lor și a calităților pe care trebuie să le îndeplinească.  
Sublinierea, oricît de puternică, a importanței activității ope-  
ratorilor de anchetă nu este nicicînd de prisos. Parafrazănd pe  
Moser, care spune că o anchetă nu poate fi mai bună decît chestio-  
narul său - ceea ce este perfect adevărat - putem formula tot  
atît de îndreptățit și ideea că o anchetă nu este mai bună decît  
lotul operatorilor folosiți. În adevăr, fără a exagera cu nimic  
lucrurile, valoarea întregii investigații depinde pînă la urmă de  
capacitatea acestor persoane de a convinge subiecții să poarte  
discuția și să o poarte în mod cît mai sincer, de conștiinciozita-  
tea lor în respectarea tuturor indicațiilor tehnice de desfășurare  
a anchetei, de rigoarea cu care înregistrează răspunsurile primi-  
te etc.

Marile institute de opinie publică din occident dispun  
de o rețea de operatori de anchetă cu care sînt capabile să acope-  
re întreg teritoriul național și să realizeze investigații într-un  
timp foarte scurt, pe baza unor comenzi și referitoare la evenimen-  
te cu derulare rapidă. Operatorii pot fi ușor mobilizați în astfel  
de campanii, deși ei nu practică o asemenea muncă decît într-un  
mod ocazional, nefiind deci vorba de o profesie. De altfel, rețeaua  
operatorilor este mereu reîmprospătată, considerîndu-se că o vechi-  
me prea mare în această activitate împieteează asupra rezultatelor.

La noi în țară, actualmente, practic, nici o instituție  
nu face apel la asemenea operatori plătiți. Se utilizează, pentru  
completarea chestionarului, diferite persoane care pot fi mobiliza-  
te, de o manieră sau alta, la acest gen de acțiuni: studenți,  
cadre didactice etc. De aici rezultă necesitatea acordării unei  
griji deosebite cel puțin pregătirii și controlului muncii unor  
astfel de operatori, dacă nu este posibilă și în selecționarea lor.  
Tot așa se poate explica și proporția relativ mare pe care o de-  
țin la noi anchetele indirecte, cu toate dezavantajele pe care a-  
cestea le prezintă; dar mulți cercetători sau colective de cerceta-  
re nu au efectiv posibilitatea de a utiliza anchetatori.

Se discută în literatura de specialitate destul de mult  
asupra calităților pe care ar trebui să le posede o persoană spre  
a fi un bun operator. Deși este evident că anumite trăsături psi-



hologice și chiar o înfățișare fizică plăcută pot avea influențe notabile asupra succesului muncii sale, pînă la urmă calitatea fundamentală a operatorului trebuie să fie totuși o trăsătură de caracter, și anume conștiinciozitatea. Orice om normal, deci neafectat de anumite defecte fizice sau intelectuale evidente, care posedă un înalt grad de conștiinciozitate, urmărind să-și ducă la bun sfîrșit munca, respectînd toate instrucțiunile primite, poate fi considerat un operator bun. Dacă această trăsătură de caracter lipsește, ea nu poate fi suplinită prin nici o altă calitate. Restul însușirilor fizice sau de personalitate pot doar influența pozitiv sau negativ, pot ridica sau scădea, într-o anumită măsură, nivelul calitativ al muncii. Mai mult, prin antrenament sau efort de voință, unele dintre trăsăturile de personalitate cu efecte negative asupra anchetei, dacă sînt conștientizate, pot fi serios modificate, cel puțin pe durata acestei munci.

Se cere, de obicei, ca operatorul să fie o persoană deschisă, extrovertită, cu facilități de comunicare cu ceilalți. Ea trebuie să fie capabilă să trezească interesul subiectului pentru anchetă și, prin urmare, să posede o capacitate de abordare a oamenilor de o manieră plăcută, destinsă etc. Prin felul său de a se purta, la început dar și pe parcursul anchetei, operatorul trebuie să se facă simpatizat de către subiect, ba chiar să-i însușească încrederea. Se cere deci ca el să fie capabil să poarte discuții pe teme diverse, să fie interesat de problemele generale ale semenilor săi, să-și poată convinge interlocutorul că-l înțelege în durerile și necazurile pe care acesta le are. Desigur, asemenea însușiri, dacă se manifestă prea puternic, pot avea un efect contrar celui dorit. Implicarea operatorului în problemele personale ale anchetatului poate contribui la influențarea lui, la crearea unei deformări constante în aprecierea informației solicitate prin chestionar.

Este evident că operatorului i se cere un anumit grad de inteligență, care-l va ajuta să se descurce în cele mai neașteptate situații, permițîndu-i adaptarea în stabilirea relațiilor cu foarte multe categorii de persoane întîlnite, diferite între ele nu numai prin nivelul cultural dar și printr-o serie de trăsături psihologice și de comportament. Nu rareori el este ținta unor atitudini agresive, de la ironiile fine ale unui intelectual pînă la



comportamentul plin de vulgarități al unui bețiv și intrînd în casă, el trebuie să ajungă să converseze cu toți acești oameni, să știe să-i găsească fiecareia "coarda sensibilă" prin care să-l atragă spre discutarea problemelor urmărite.

Se mai apreciază că o calitate necesară a oricărui bun operator de anchetă este exactitatea, rigoarea. Intr-adevăr, un minim spirit de rigoare nu trebuie să lipsească în munca nici unui anchetator, deoarece altfel este de așteptat apariția unui număr exagerat de mare de erori în înregistrarea informației. Dar consider că această însușire - ca și cele menționate mai sus - poate fi dobîndită printr-un antrenament adecvat și aproape oricine poate ajunge să completeze exact un chestionar.

Referitor la nivelul de instrucție al operatorilor, specialiștii înclină spre a-i considera mai potriviți pe cei cu un nivel mediu, dat fiind că persoanele cu un standard cultural foarte ridicat ajung să se plictisească mai repede de operațiile, la urma urmei de rutină, pe care le presupune completarea unui număr mare de chestionare. Desigur, această apreciere poate fi discutată; dacă operatorul manifestă un interes oarecare în cercetare, dacă el este stimulat să participe la investigație și din alt punct de vedere decît ca simplu operator, atunci nivelul de cultură și chiar unele cunoștințe de specialitate pot să fie de bun augur pentru cercetare. Am ajuns astfel să ating și această problemă delicată a interesului operatorului pentru cercetare. Ea este realmente o problemă, mai ales la noi unde, cum spuneam, munca operatorilor nu este remunerată; fără un minim de interes din partea celor angajați la realizarea anchetei, nu este de conceput o cercetare reușită. Fie că este vorba de studenți, de cadre didactice sau de alte persoane folosite ca operatori, ele trebuie, pe de o parte, să înțeleagă importanța, științifică sau practică, a studiului și, pe de altă, să fie implicate în cît mai mare măsură și în alte faze ale cercetării.

Rezultă deci, din cele spuse pînă aici, că, și în cazurile în care sociologul nu are prea multe variante de alegere a echipei de operatori, se cer cîteva lucruri întreprinse pentru a nu ajunge la un eșec complet al cercetării.

În primul rînd, trebuie eliminate din lotul potențial de operatori acele persoane despre care se știe de la bun început



că au grave curențe de ordin moral, fizic sau de altă natură, persoane care prin prezența lor în câmpul cercetării ar putea constitui o indubitabilă sursă de erori.

În al doilea rând, să se asigure o instruire completă și complexă a operatorilor cu care se va lucra, prin care să se urmărească obiectivele următoare:

(i) Cunoașterea exactă a scopului și importanței cercetării. La acest capitol se vor accentua efectele și beneficiile pentru societate sau pentru dezvoltarea sociologiei, contribuind astfel la stimularea motivației participării operatorilor la buna desfășurare a anchetelor. Se va insista asupra răspunderii ce revine întregii echipe de cercetare pentru obținerea unei informații corecte, nedistorsionate, lucru ușor de realizat atunci când rezultatele finale vor fi puse la dispoziția organelor de conducere, constituind o bază pentru luarea deciziilor. De asemenea, se pot preciza, așa cum am mai spus, eventualele posibilități de participare a celor doritori la fazele de prelucrare și interpretare a datelor.

(ii) Familiarizarea operatorilor cu chestionarul ce urmează a fi aplicat. Fiecare dintre aceștia trebuie să cunoască sensul exact al tuturor întrebărilor, felul în care acestea se adresează, precum și modalitatea concretă de înregistrare a datelor. Este recomandabil ca în prezența întregului lot de operatori să se realizeze completarea câtorva chestionare, în situații de simulare. (Unul dintre sociologii cu multă experiență să joace rolul subiectului imaginar iar câțiva dintre operatori să încerce pe rând să-l chestioneze). Se vor evidenția cu această ocazie toate posibilitățile permise operatorilor în a schimba eventual ordinea întrebărilor în a da explicații suplimentare atunci când o întrebare nu este corect înțeleasă (eventualele libertăți în reformularea textului întrebărilor) etc.

(iii) Precizarea modalităților de găsim și identificare a persoanelor de chestionat și alternativele pentru cazul când persoanele în cauză nu pot fi găsite, refuză sau, din alte motive, se află în imposibilitatea de a răspunde la întrebări. Se vor transmite cu această ocazie informații asupra modului de alegere a eșantio-



nului, cât și unele sumare cunoștințe legate de noțiunea de reprezentativitatea a acestuia. Astfel, operatorii vor fi puși în gardă că orice abatere de la regula de alegere a indivizilor stabilită constituie o sursă de eroare, care poate anula multe din eforturile lor de a conduce cât mai exact ancheta.

(iv) Specificarea posibilităților concrete de rezolvare a tuturor problemelor ce se pot ivi pe teren, legate de aplicarea chestionarului și, în general, de activitatea lor de investigații. Se va atrage cu deosebire atenția că toate problemele ce apar trebuie rezolvate de o manieră uniformă, în același mod pentru toți operatorii, și că inovațiile personale ale unuia sau altuia nu vor avea alt efect decât să conducă la inutilizarea chestionarelor sau, în cel mai bun caz, a unor părți din chestionarele astfel modificate.

În al treilea rând, este necesar să se asigure un control cât mai riguros posibil al muncii de teren. Operatorii nu pot fi, decât în cazuri cu totul excepționale, distribuiți direct în teren și reîntâlniți doar în finalul cercetării, ci trebuie urmăriți pas cu pas în activitatea lor. E vorba, desigur, mai ales de a verifica primele chestionare efectuate de fiecare, spre a se elimina din timp greșelile inerente ce apar din cauza neînțelegerii corecte a sensului unor întrebări, a modalităților de înregistrare a informației sau datorită altor factori. Apoi, dacă lucrul este posibil, se cere verificat, prin sondaj, dacă adresele primite de fiecare au fost într-adevăr respectate și eventual chiar repetarea unor interviuri de către sociolog. Este limpede că sarcinile controlului se vor reduce substanțial dacă recomandările precedente au fost cât mai riguros realizate. De asemenea, este evident că orice operator surprins că se pretează la fraudă trebuie eliminat din echipă; la fel și cei care, din diferite alte motive, aduc chestionare cu un număr nepermis de mare de erori.

Având în vedere că, în multe cazuri, lotul operatorilor este constituit din persoane care-și oferă voluntar, sau aproape, serviciile se cere ca întreaga activitate de control să se facă de o manieră care să nu lezeze personalitatea acestora, să nu jignească pe nimeni. Asta nu trebuie să ducă însă la trecerea cu vedere a nici unei nereguli. E nevoie deci de o îmbinare optimă a fer-



mității cu spiritul de politețe și respect față de cei cu care sociologul este pus în situația să colaboreze într-o cercetare.

## 2. Chestionarul

Am precizat de la început că instrumentul cu ajutorul căruia se realizează culegerea datelor în achete este chestionarul și, prin urmare, va fi nevoie să ne oprim puțin asupra lui, spre a pune în evidență unele principii care stau la baza întocmirii sale, formele pe care le poate lua în diferite tipuri de anchetă.

Definiția simplă de la care am plecat - chestionarul ca o succesiune de întrebări - are nevoie de câteva adăugiri și comentarii, deoarece este evident, pentru oricine că nu orice set de întrebări poate primi numele de chestionar (sociologic). Dar nu doresc nici să caut o definiție propriu-zisă, care să reziste unei analize serioase, căci mă tem că nu este posibil așa ceva. Ceea ce mi se pare totuși posibil și necesar este faptul de a distinge chestionarul sociologic de alte forme asemănătoare, cu mare cîmp de circulație în viața de fiecare zi. Am spus chestionarul sociologic și prin asta am introdus deja un element de delimitare, rezultat din însăși finalitatea pentru care a fost construit instrumentul respectiv. Am senzația că oricît ne-am strădui, fără această notă, din care să rezulte scopul în care va fi folosit, este pînă la urmă imposibil să distingem un chestionar sociologic de o serie de documente prin care se transmite informația, de exemplu, între diferite niveluri ierarhice ale unei organizații sau de alte instrumente asemănătoare.

Prin urmare, chestionarul sociologic va cuprinde o serie de întrebări ce se adresează oamenilor care formează populația supusă investigației, întrebări ce joacă rolul de indicatori ai unor fapte sau fenomene sociale. Răspunsul primit la aceste întrebări - înregistrat de regulă în scris - furnizează o informație asupra cîmpului problematic cercetat, informație al cărei statut depinde de natura întrebărilor puse. Pentru precizarea acestui aspect este normal să reluăm clasificările tradiționale ce se folosesc pentru întrebările chestionarelor.



I Mai întâi, după conținut, întrebările se pot împărți în trei mari categorii: factuale, de opinie și de cunoștințe.

Ia) Așa cum o arată și numele, întrebările factuale vizează obținerea unei informații legate de anumite evenimente petrecute în viața persoanei, a familiei sale, de unele însușiri obiective ce o caracterizează etc., deci de fapte, de stări care există așa cum există, independent de interpretările, opiniile pe care le poate avea omul în legătură cu ele. Numărul de copii, anul nașterii, retribuiția din ultima lună, ce emisiuni a urmărit ieri la televizor, dacă are apă curentă în casă etc. - iată doar câteva exemple de informații ce se urmăresc a se obține prin întrebările factuale. O caracteristică a acestor întrebări factuale o constituie posibilitatea - cel puțin în principiu - a verificării exactității răspunsurilor primite prin alte mijloace, cum sînt documentele, observația directă, chestionarea altor persoane etc.

IIb) Prin întrebările de opinie - și aici termenul de opinie trebuie luat într-un sens foarte larg - se urmărește accesul la aspectele de ordin subiectiv ale persoanei: atitudini, păreri, motive, interese etc. Tot acest univers psihologic al individului prezintă un interes deosebit pentru sociolog (și nu numai pentru cercetătorii din domeniul psihologiei), deoarece aspectele de conștiință nu rămîn izolate în sfera individuală, ci ele au un rol foarte important asupra comportamentului și acțiunilor oamenilor, fiind deci strîns legate de faptele propriu-zise. Mai mult, anumite concepții, păreri, atitudini etc., în cadrul social, se desprind de purtătorii individuali, deși ele nu există decît prin aceștia, devenind fapte sociale de sine stătătoare. Prin urmare, anchetele îndreptate în această direcție nu trebuie privite numai - sau în primul rînd - ca încercări de a descrie stările psihice individuale, ci unele elemente ale conștiinței colective, ca probleme detașabile, într-un anume sens, de indivizii de la care pornesc și prin care subzistă. Opiniile și atitudinile oamenilor în legătură cu munca sau cu alte aspecte ale vieții se concretizează în valori și comportamente sociale, legate de aceste elemente, va în valori și comportamente care la rîndul lor acționează asupra opiniilor și atitudinilor individuale.

Așadar, ne aflăm în fața unui univers problematic deosebit de complex, care ridică în fața cercetătorului multe obstacole



nu numai în abordarea dar și în interpretarea fenomenelor studiate. Cît privește obstacolele metodice, ele se impun cu evidență oricărui cercetător și se relevă mai simplu dacă vom compara situația întrebărilor de opinie cu cea a întrebărilor factuale. Mai întâi, despre anumite fapte știm dacă este sau nu oportun să punem oamenilor întrebări, adică știm dacă metoda anchetei poate duce la o informație consistentă. Îl voi putea întreba liniștit pe fiecare om în legătură cu nivelul școlar pe care l-a dobândit, plecînd de la premisa normală că fiecare știe cîte clase a urmat. În general universul întrebărilor factuale poate fi lărgit la toate aspectele despre care sîntem convinși că individul uman deține informația de care avem nevoie. Această delimitare este relativ facilă, singura problemă ce poate interveni fiind aceea dacă omul vrea să ne răspundă corect, adică se cer evitate acele întrebări pentru care există bănuiala că oamenii, dintr-un motiv sau altul (mai cu seamă de teama unor repercusiuni negative asupra vieții lor), nu vor răspunde.

În schimb, în legătură cu atitudinile, opiniile etc. ar trebui aproape de fiecare dată să ne punem în mod serios problema dacă ele există, dacă are sau nu sens să adresăm oamenilor asemenea întrebări. De multe ori nu se ține cont de acest lucru, considerîndu-se de la sine înțeles că putem chestiona o persoană despre, să zicem, satisfacția pe care o are în muncă, părerea sau atitudinea față de un anumit eveniment social important și astfel o parte însemnată din răspunsuri nu vor avea nici o acoperire, ele fiind date din complezență sau din teama omului că ar fi ridicol dacă ar recunoaște cîștit că nu are nici o părere în legătură cu ceea ce-l întrebăm.

Un al doilea gen de dificultăți metodologice legate de întrebările de opinie derivă din caracterul extrem de imprecis al problemei astfel urmărite, ceea ce face ca orice schimbare nu numai în textul întrebării dar chiar și în tonalitatea cu care ea este adresată să producă modificări de accente și deci de răspunsuri. Este motivul pentru care aproape în toate anchetele uzînd de astfel de întrebări se fac precizări exacte asupra modului cum se adresează întrebările de opinie și se interzice operatorului orice schimbare, orice interpretare personală a formulării dată în



chestionar. Dacă la întrebările factuale operatorul are posibilitatea să uzeze, de cele mai multe ori, de orice intervenție ce-o crede potrivită pentru a stimula subiectul să răspundă sau să găsească răspunsul corect, aici situația este complet diferită. Tot întrebările de opinie sînt cele care determină recurgerea la anchete directe pentru a nu da astfel timp de gîndire subiectului și posibilitatea de a inventa sau căuta în jur opinii ce nu-i aparțin.

pro  
nec  
In sfîrșit, a treia problemă metodologică de bază legată de întrebările de opinie este ridicată de încercarea cercetătorului de a măsura intensitatea anumitor opinii, în speță a atitudinilor. Psihologia socială a pus la punct un întreg arsenal în acest scop, dar spațiul și profilul cursului de față nu-mi permit să intru în detaliile acestei probleme complexe.

Înainte de a încheia discuția despre întrebările de opinie aș dori să mai fac o precizare importantă, și anume că opiniile oamenilor nu trebuie niciodată luate drept altceva decît opinii. Nu de puține ori sociologul încearcă explicații sau soluții la diverse probleme sociale și atunci efectuează o anchetă de opinie spre a vedea care sînt părerile oamenilor legate de aspectele respective, avînd impresia că astfel va ajunge la ceea ce urmărește el. Nu este exclus ca din analiza acestor opinii să poată găsi explicația sau soluția, dar, oricare ar fi situația, baza explicativă sau de rezolvare a problemei nu se găsește în opiniile exprimate. Chiar dacă oamenii cred în proporție de 100% că un fenomen social negativ, să zicem, se datorează unei cauze anumite, sociologul nu are dreptul să se mărginească la o astfel de explicație. Aici, ca și în multe alte probleme asemănătoare, ceea ce află el nu este decît interpretarea pe care o dau oamenii lucrurilor, interpretare ce poate fi corectă sau falsă și care ea însăși se cere interpretată și explicată. Sociologul își pierde, evident, statutul de om de știință - la care știm că aspiră atît de puternic - dacă transferă indivizilor din colectivitatea studiată responsabilitatea găsirii soluției problemelor sociale, iar sociologia s-ar transforma în acest caz într-o disciplină care ar culege explicații în loc să le furnizeze ea însăși. Am dat exemplul explicației și cel al rezolvării problemelor sociale pentru că sînt cele mai elocvente în a susține ideea de mai sus; dar ele nu sînt singurele chestiuni în





care opiniile trebuie tratate doar ca opinii, problema fiind, în același timp, generală dar și specifică, prin nunațe, fiecărui aspect studiat.

1c) Al treilea gen de întrebări, în funcție de conținutul lor, îl reprezintă cele de cunoștințe. În acest caz, informația obținută de sociolog nu este legată de conținutul întrebării, ci de faptul dacă omul deține sau nu anumite cunoștințe. Dacă-l întreb la ce oră s-a sculat în dimineața asta, direct să aflu efectiv acest lucru, tot așa cum direct să-i evaluez, de pildă, gradul de satisfacție cu locuința în care trăiește. În schimb, dacă-l întreb cine a scris romanul "Ion" este clar că nu mă va interesa informația ca atare, ci coincidența răspunsului cu unul standard, considerat corect - "Rebreanu", în cazul de față - sau noncoincidența.

Așa cum bine se înțelege, rolul întrebărilor de cunoștințe este, de regulă, acela de a evalua, prin câțiva indicatori foarte expresivi, nivelul de cultură (de fapt, al unor dimensiuni ale culturii) specific diferitelor grupuri umane. Adiacent, pot apare de aici și informații despre sistemul de valori, atitudini etc., dacă interpretăm prezența sau absența cunoștințelor în anumite domenii. Problemele metodologice cele mai importante ridicate de întrebările de cunoștințe se referă la alegerea unor indicatori capabili să reflecte realmente nivelul în zona culturală investigată și apoi, mai pragmatic, de a găsi modalitatea cea mai potrivită de a adresa asemenea întrebări. Este evident că plăcerea omului de a răspunde la un chestionar, despre care am pomenit, dispare aproape complet atunci când el se simte "verificat" prin întrebări de cunoștințe. Din această cauză, se recomandă multă, foarte multă prudență în utilizarea acestor întrebări.

În rândurile de mai sus am preluat o clasificare clasică a întrebărilor după conținut. Se poate observa că în unele manuale se vorbește despre o clasificare a chestionarelor după același criteriu. Consider însă că prea rare sînt situațiile când avem de a face cu chestionare "pure", formate numai din întrebări factuale, numai de opinie sau numai de conținut. De regulă, chestionarul cuprins de mai multe genuri de întrebări; în speță sînt aproape inexistente chestionarele formate doar din întrebări de opinie, acestora alăturîndu-li-se cel puțin cîteva întrebări factuale, cu



rol de clasificare a indivizilor după vîrstă, sex, nivel școlar etc. Aceeași observație este valabilă și pentru chestionarele axate precumpănitor pe întrebări de cunoștințe. Chestionarele pur factuale pot avea ceva mai multe șanse de apariție, dar și aici, de obicei, sociologul este tentat să strecoare cîteva întrebări de opinie care să întregească universul informațional cu elemente de apreciere, mulțumire etc. sau pur și simplu pentru a introduce unele aspecte capabile să rupă monotonia unei anchete bazată numai pe date factuale.

II A doua clasificare clasică a întrebărilor este cea privind forma de înregistrare a răspunsurilor. Se vorbește, în acest caz, de întrebări deschise, închise și mixte (sau semiînchise ori semi-deschise).

II a) Întrebările deschise sînt cele urmate în chestionar de un spațiu liber în care se va înregistra răspunsul. În ancheta indirectă, subiectul fiind cel care completează chestionarul, va umple așa cum crede el de cuvință acest spațiu, adică va putea realiza fie un comentariu larg al problemei fie un rezumat scurt de cîteva cuvinte sau chiar de unul singur. Cînd se lucrează cu operatori de anchetă, aceștia sînt, de regulă, instruiți să noteze răspunsul subiectului în mod cît mai detaliat și mai fidel, interzîcîndu-se deci rezumarea sau interpretarea personală din partea operatorului. Desigur că o asemenea cerință rămîne de cele mai multe ori un deziderat greu de împlinit și chiat nerealizabil, avînd în vedere multitudinea situațiilor întîlnite pe teren. În orice caz, se pare că cel puțin o rezumare a răspunsurilor ample este inevitabilă, operatorul trebuînd să se străduiască să nu pierdă nici una din ideile principale și să nu le altereze sensul.

II b) Întrebările închise sînt cele însoțite de variante posibile de răspuns, subiectul sau operatorul fiind obligați să aleagă pe cea potrivită. Spun subiectul sau operatorul deoarece se pot ivi ambele situații. Subiectul este chemat să aleagă varianta de răspuns întotdeauna în ancheta indirectă și uneori în cea directă, atunci cînd după citirea întrebării operatorul are sarcina să anunțe și variantele și să roage subiectul să aleagă varianta în conformitate cu opinia sau faptul în cauză. Alteori în ancheta directă subiectului nu i se aduc la cunoștință variantele de răspuns, între-



barea avînd pentru acesta formă deschisă, dar operatorul încadrează răspunsul în una din variantele prevăzute.

Iată cîteva exemple de întrebări închise:

A. Care este vîrsta dumneavoastră ?

- |                |                    |
|----------------|--------------------|
| 1. Sub 20 ani  | 5. 50 - 59 ani     |
| 2. 20 - 29 ani | 6. 60 - 69 ani     |
| 3. 30 - 39 ani | 7. 70 ani și peste |
| 4. 40 - 49 ani |                    |

B. În ce măsură sînteți satisfăcuți de numărul de camere din locuința dv. ?

1. Foarte mult
2. Mult
3. Moderat
4. Puțin
5. Foarte puțin

C. Care sînt modalitățile preferate de dv. pentru petrecerea timpului liber ?

1. Lectură
2. Cinematograf
3. Televizor
4. Excursii, plimbări
5. Practicarea sportului
6. Asistență la manifestări sportive
7. Dans
8. Teatru

9. Opera

Răspunsul la întrebările închise se realizează prin marcarea variantei de răspuns aleasă în diferite forme: fie încercuind cifra de cod care, de regulă, precede varianta respectivă, fie trecînd aceeași cifră într-o căsuță (pătrățel) construit alături de întrebare, pe marginea foii, spre a leși mai clar în evidență, fie prin alte metode (sbuliniere, marcarea cu X etc.) Variantele de răspuns pot fi sau nu exclusive. Astfel, în primele două exemple de întrebări este clar că nu se poate marca decît o variantă, pe cînd în a treia se pot alege două, trei sau chiar mai multe.

Una dintre problemel tehnice cele mai importante legate



de construcția întrebărilor închise este cea care privește epuizarea întregului univers referențial prin variantele alese, adică cercetătorul trebuie să prevadă, din etapa proiectării investigației, toate posibilitățile de răspuns. Din cele trei exemple alese, primul nu ridică nici o problemă de acest gen. Al doilea, aparent și el surprinde, prin forma de scală ordinală, un eventual satisfăcător de variante; o singură chestiune poate apărea în cazuri ca acesta, și anume imposibilitatea subiectului de a emite opinia cerută. De aceea, pentru a nu forța subiectul să dea un răspuns neconform cu realitatea, este de multe ori preferabil ca acestor variante de răspuns să li se mai adauge una din genul: "nu știu", "nu mă pot pronunța", "nu m-am gândit la problema aceasta" etc.

Intrebarea C am pus-o intenționat într-o formă defectuoasă din punctul de vedere al surprinderii tuturor posibilităților de răspuns. Este evident că lista propusă acolo trebuia continuată cu încă alte variante de petrecere a timpului liber, dar tot atât de adevărat este că la întrebarea noastră pot apărea cele mai neașteptate răspunsuri și, prin urmare, e greu de a respecta cerința exhaustivității chiar dacă lista s-ar îmbogăți substanțial. În aceste condiții, se recurge la cea de a treia formă de întrebări - mixte sau semiînchise (semideschise) - ce cuprind, pe lângă setul de variante explicitate, încă una din genul "Alte situații. Care? ... " Se cere astfel atât încercuirea unui cod - atașat acestei variante - dar și exprimarea liberă a situației neprevăzută în variantele scrise.

Cît privește frecvența de utilizare a celor trei tipuri de forme de înregistrare a răspunsurilor, se poate emite următoarea apreciere. Întrebările factuale sînt aproape în totalitate redactate sub formă închisă, întrucît variantele de răspuns sînt mai ușor previzibile. Doar atunci cînd numărul alternativelor este prea mare, după însiruirea cîtorva, socotite principale (la care se așteaptă frecvența de răspuns mai ridicate) se poate apela la varianta "Alte situații". De pildă, la întrebările obișnuite despre ocupație, putem folosi următoarea schemă:

1. agricultor
2. muncitor
3. maestru
4. tehnician, proiectant, subinginer



5. funcționar
6. lucrător în servicii
7. intelectual
8. alta . . . . .

În cazul întrebărilor de opinie, distingem în special clasa aceloră în care se urmărește evaluarea intensității unei atitudini, aprecieri, atașament la o valoare etc., caz în care se utilizează pentru răspuns o scală ordinală închisă, ca la întrebarea B. Treptele scalei pot fi, cum sînt acolo, în număr de 5, mai mic sau mai mare.

Iată încă două exemple:

D. Cum apreciați emisiunea X, din programul televiziunii din ziua.

?

1. Bună
2. Mediocră
3. Slabă

E. În ce măsură sînteți mulțumiți de modul de gospodărire a localității în care trăiți ?

1. Foarte mult
2. Mult
3. Relativ mult
4. Nici mult, nici puțin
5. Relativ puțin
6. Puțin
7. Foarte puțin

Evident, ordinea variantelor poate fi cea dată aici sau una inversă, începînd deci cu aprecierile negative. Această din urmă situație se întâlnește mai ales atunci cînd numărul din fața variantei este luat nu pur și simplu ca simbol de cod, ci drept măsură a intensității opiniei respective. La limită, în asemenea cazuri se poate ajunge să nu se mai treacă deloc variantele de răspuns pe chestionar, ci punînd de exemplu întrebarea E, să se precizeze: "dați o notă de la 1 la 10, care să reflecte opinia dv., convenind că 1 corespunde insatisfacției totale iar 10 satisfacției maxime".

Restul tipurilor de întrebări de opinie - care nu se pretea-



ză la o astfel de scalare a răspunsurilor -, exceptând cele care cer răspunsuri în cadrul dihotomiei "Da/Nu" și care, la limită, pot fi considerate ca aparținând tipului de mai sus, restul întrebărilor deci pot fi mai greu surprinse sub forma evantaiului închis de răspunsuri și ele vor apare, în bună măsură, sub formă deschisă sau, eventual, semiînchisă.

În ceea ce privește întrebările de cunoștințe, ele se pot prezenta în egală măsură în cele două tipuri formale principale. Deși aparent forma deschisă pare cea mai potrivită pentru ele, adesea se folosește un evantai de răspunsuri prestabilite, cerându-i-se subiectului să aleagă varianta corectă (așa cum se procedează și în chestionarele folosite la examenul pentru obținerea carnetului de conducere auto). Această modalitate din urmă este utilizată cu deosebire în ancheta indirectă (autocompletare) pentru a evita amestecul deprinderilor de scris și cunoștințelor de limbă cu cele urmărite de cercetare. Un exemplu de astfel de întrebare poate fi:

F. Cine este autorul romanului "Ion" ?

1. Sadoveanu
2. Agîrbiceanu
3. Rebreanu

Recapitulînd și încercînd o evaluare globală, putem constata că întrebările închise sînt cele care predomină de departe în marea majoritate a chestionarelor sociologice. Rațiunile pentru care se întîmplă acest lucru ar putea fi sistematizate astfel:

Mai întîi - și cu siguranță, că aceasta este motivul fundamental - la egalitatea celorlalți factori, întrebările închise sînt preferate pentru facilitatea prelucrării informației obținute. Orice formă ar lua răspunsurile la o întrebare, pînă la urmă ele trebuie reduse și aduse la niște elemente standardizate care să ofere posibilitatea comparării informației obținute de la diferiți indivizi, ceea ce înseamnă că și întrebările libere trebuie, în ultimă instanță "închise" într-un sistem clasificatoriu. Or, asemenea muncă este extrem de migăloasă și, așa cum au arătat unele experimente metodologice, ea este acompaniată de un procent ridicat de erori de codificare. În esență, prelucrarea întrebărilor deschise este o veritabilă analiză de conținut și se înțelege că



dacă numărul acestor întrebări este mare, efortul solicitat de sistematizarea informației e atât de ridicat încât poate depăși cu ușurință capacitățile echipei de cercetători. La întrebările închise, rubricile de prelucrare sînt gata elaborate astfel că adesea informația se trece direct de pe chestionar în sistemul automat de prelucrare a datelor. Întrebările semiînchise pot ridica unele probleme la prelucrare dacă frecvența răspunsurilor la varianta liberă (nedeterminată) este mare, ceea ce înseamnă că s-au omis unele aspecte ce nu pot fi neglijate și este nevoie deci ca ele să fie puse în evidență prin analiza răspunsurilor concrete.

Al doilea motiv ce poate fi invocat în favoarea întrebărilor închise este legat de facilitatea mînuirii unui chestionar în care predomină întrebările de acest gen, deci de posibilitățile rapide de completare, fapt ce-i asigură o eficiență practică deosebită.

În al treilea rînd, întrebările închise prezintă și un avantaj legat de ajutorul pe care variantele de răspuns îl dau subiectului în înțelegerea corectă a sensului întrebării din chestionar. Acest lucru este valabil mai ales în anchetele indirecte, cînd anumite defecțiuni în formularea întrebării sau ambiguitatea unor termeni pot conduce la interpretarea neuniformă a unei întrebări. Or, văzînd care sînt posibilitățile de răspuns, subiectul este într-un fel dirijat pe făgașul urmărit de anchetator.

În sfîrșit, o problemă apropiată de precedenta, dar nu identică, ar fi aceea că variantele de răspuns pot "provoca" o anumită opinie sau reactualiza o informație altfel ignorată. De multe ori omul nu are clar conștientizată o anumită atitudine sau opinie, pe care însă subiectul o regăsește imediat ca fiind a lui atunci cînd citește sau i se citesc variantele unei întrebări. Ideea merge, cum spuneam, și la întrebările factuale. De exemplu, dacă cerem omului data unui eveniment mai îndepărtat în timp el nu va fi capabil s-o spună cu exactitate; în schimb, găsind niște variante de răspuns construite pe intervale de timp delimitate de anumite evenimente istorice, el nu va avea nici o dificultate să încadreze corect evenimentul.

Cu acest aspect al problemei ajungem și la dezavantajele




întrebărilor închise. Așa cum mai spuneam, omul nu are întotdeauna opinii formate și o întrebare închisă îl poate ajuta să iasă cu fața curată, indicînd la împlinire una dintre variante. Alte persoane, dimpotrivă, se vor simți stînjinite de unele delimitări categorice așa cum apar prin variantele indicate, considerînd că un răspuns liber ar fi mai adecvat, le-ar permite să-și exprime mai nuanțat părerile. De asemenea, un chestionar bazat aproape în exclusivitate pe întrebări închise poate crea o atmosferă de "răceală" între operator și subiect, de rigiditate sau chiar de superficialitate în răspunsuri.

Unele dintre neajunsurile întrebărilor închise sînt eliminate în ancheta directă, atunci cînd operatorul nu aduce la cunoștință subiectului variantele de răspuns, ci el este cel care o alege pe cea potrivită în urma unui răspuns liber. În orice caz, întrebările închise s-au statornicit cu multă forță în chestionarele sociologice și este evident că acest lucru s-a întîmplat tocmai datorită avantajelor lor incontestabile. Întrebările libere punctează din loc în loc chestionarul pentru a face discuția mai atractivă, pentru a stimula subiectul cu anumite probleme deosebite, și desigur, pentru a căuta nuanțe și informații suplimentare la acele chestiuni importante ce nu pot fi expediate în cîteva sumare variante de răspuns.

III O altă clasificare a întrebărilor, ce se cuvine amintită aici, este cea care distinge întrebările directe de cele indirecte. În adevăr, dat fiind faptul că omul nu este întotdeauna dispus, din motive extrem de diverse, să dezvăluie o anghmită informație pe care cercetătorul o așteaptă, acesta din urmă este pus în situația de a face cîteodată apel la unele căi ocolitoare de atingere a scopului urmărit. O întrebare va fi numită indirectă dacă sensul real al său sau al răspunsului este diferit de cel aparent, adică dacă prin intermediul său se urmărește un lucru care nu este sesizat de cel interogată. Am amintit mai înainte că prin unele întrebări de cunoștințe putem urmări nu atît ce și cît cunoaște un om, ci profilul unui gen de cunoștințe spre a ne dezvălui eventuale orientări valorice. Există multe genuri de probleme sociologice care pot fi studiate doar prin astfel de întrebări indirecte, cum ar fi, de pildă, gradul de religiozitate, starea materială, comportamentul sexual etc. Este limpede că nu vom adresa



oamenilor direct întrebări de tipul: "Credeti în Dumnezeu?", "Câți bani câștigați pe lună în afara retribuției?" sau altele la care răspunsurile vor fi evident distorsionate de teama anumitor consecințe negative, de tendința indivizilor de a se conforma - formal - unor norme și valori sociale dominante, în general de sistemul de autoapărare pe care fiecare și-l formează în relațiile cu semenii săi.

La încheierea acestui paragraf, doresc să mai spun doar  câteva cuvinte despre numărul și ordinea întrebărilor din chestionar. Mai întâi o precizare: în ciuda impresiei care poate persista în rîndul unor nespecialiști, cum că atîtea decenii de cercetare sociologică empirică ar fi condus la găsirea unor soluții optimale, general valabile, în legătură cu cele două caracteristici ale chestionarului menționate, nu există totuși o rețetă cu aplicabilitate universală nici în privința numărului și nici a ordinii întrebărilor.

Chestionarele pot conține de la una-două întrebări la 300 sau chiar mai multe și e imposibil de spus dinainte ce număr este preferabil. Acesta ~~depinde~~ de o mulțime de factori care țin de conținutul problemei de studiat, de complexitatea faptelor și fenomenelor urmărite, de bogăția interacțiunilor cu altele, mai țin de cunoștințele pe care cercetătorul le are despre aceste fenomene, de consistența schemelor teoretice și explicative de care dispune, dar și nu în ultimul rînd - de aspecte exterioare, în speță de mijloacele materiale și timpul avut la dispoziție. În plus, numărul întrebărilor va depinde de tehnica de aplicare a chestionarului (o anchetă prin poștă va face apel la chestionar mai scurt decît una directă) și de populația căreia i se adresează (în general, persoanele cu studii superioare sînt mai ușor de antrenat în discuții pe baza unor chestionare lungi sau le pot completa și singure cu mai multă ușurință).

Referitor la ordinea întrebărilor, se poate aprecia că aceasta capătă o importanță deosebită în următoarele două cazuri:

a) cînd conținutul unor întrebări este de așa natură încît ele trebuie așezate la distanță cît mai mare una de alta pentru ca răspunsurile să nu se influențeze reciproc și



b) cînd chestionarul conține unele întrebări "dificile" care pot trezi neîncrederea subiectului sau genera refuzul de a mai răspunde; astfel de întrebări se vor așeza, evident, spre sfîrșitul chestionarului pentru a diminua pe cît posibil efectele negative.

Tot în finalul chestionarelor în care predomină întrebările de opinie se recomandă - mai ales dacă subiectului i se garantează anonimatul - plasarea întrebărilor factuale de clasificare, întrebări care uneori pot lăsa impresia că prin ele se urmărește culegerea cît mai multor indici pentru identificarea ulterioară a persoanei care răspunde. Este bine ca, ajunși în acest punct cu interviul, operatorii să lămurească exact rostul întrebărilor de clasificare în chestionar.

### 3. Erori și surse de erori în anchete

Datorită faptului că ancheta este o metodă de investiga-re în cursul desfășurării căreia intervin nenumărați factori perturbatori, manualele de metode și tehnici se străduiesc, în mod cu totul îndreptățit, să inventarieze cît mai complet sursele care provoacă erori, în diferitele faze ale cercetării. O idee de principiu ce se cere subliniată de la început este aceea că, de regulă, o eroare nu este generată de un singur factor, ci apare ca rezultat al interacțiunii tuturor factorilor ce intervin la momentul respectiv al cercetării. Cu toate acestea, obiceiul este să se treacă în revistă erorile după factorul principal generator, ceea ce voi face și eu în continuare, dar ideea de mai sus trebuie să rămînă mereu prezentă în minte spre a evita riscul unor interpretări unilaterale.

#### 3.1. Erori generate de construcția chestionarului

Deja din cele spuse pînă aici s-au degajat cîteva aspecte de construcție a chestionarului generatoare de erori, cum ar fi conținutul întrebărilor, lungimea excesivă a chestionarului sau ordinea nepotrivită a întrebărilor, aspecte asupra cărora nu voi mai reveni în continuare.



O problemă, implicînd de fapt ordinea întrebărilor, dar cu o semnificație mult mai largă și cu un rol foarte important în producerea erorilor, mai cu seamă dacă e vorba de anchete indirecte, este structura (logică și grafică) a chestionarului. Mă gîndesc aici în special la faptul că un chestionar este rareori constituit din întrebări care toate se adresează întregului lot anchetat. Adică, altfel spus, apar frecvent întrebări ce nu se vor adresa anumitor persoane, pentru că ele nu vizează situații specifice acestora. De exemplu, dacă studiem calitatea vieții într-o colectivitate, vom ajunge inevitabil să abordăm și aspecte legate de muncă, de locul de muncă, de transportul spre acesta etc. Or, este evident că o parte a populației, fiind inactivă, nu va fi vizată de asemenea întrebări. În toate aceste cazuri, chestionarul trebuie, prin mențiuni foarte clare și vizibile, să pună în evidență ce categorii de persoane răspunde la o anumită întrebare. Apar astfel în chestionare așa-numitele întrebări filtru care orientează ★ subiectul spre întrebările la care să răspundă. În exemplul de mai sus, o asemenea întrebare filtru, avînd, să zicem numărul 30, va fi următoarea:

30. Sînteți încadrat în cîmpul muncii ?

1. Da

2. Nu

Presupunînd că după ea urmează 10 întrebări referitoare la persoanele active (numerotate de la 31 la 40), va fi necesar ca întrebarea filtru, cu nr.30, să fie însoțită de mențiunea: "Cei care răspund 'da', vor continua cu întrebarea 31, iar cei care răspund 'nu', vor trece direct la întrebarea 41". Sau - și mai bine - întrebările 31-40 vor fi puse într-un chenar deasupra sau în dreptul căruia se va specifica în mod foarte vizibil: "Numai pentru persoanele care au răspuns 'Da' la întrebarea 30, deci cele încadrate în cîmpul muncii".

Desigur, sînt și alte modalități de vizualizare a acestor bifurcații și nu este niciodată de prisos să se ia cît mai multe măsuri de precauție, chiar și atunci cînd se lucrează cu operatori de interviu, instruiți în prealabil. Erorile de această natură sînt practic prezente în orice anchetă indirectă căci este aproape imposibil ca oamenii să nu completeze și întrebări care nu-i pri-



la interpretări eronate ale răspunsurilor subiecților, care vor fi scrise ca atare, la întrebările libere sau traduse în coduri, la cele închise. O sursă interesantă de erori, pusă în evidență de unele studii metodologice - vezi Moser (1967, pp. 356-357) - o reprezintă așa numitele anticipații ale operatorilor de anchetă.

Operatorul poate fi influențat în completarea chestionarului de răspunderile pe care le primește la primele întrebări, de unde el își face o imagine asupra atitudinii de ansamblu a subiectului în legătură cu problema studiată, avînd deci tendința să interpreteze orice răspuns în lumina acestei imagini-model (anticipațiile de structură-atitudine). Pe de altă parte, luînd contact cu persoana de anchetat, operatorul observă unele elemente ce țin de vîrstă, condiția socială și intelectuală a ei, de starea materială etc. și deci situează această persoană pe o anumită treaptă socială, o identifică cu un anumit tip social. În consecință, orice ambiguitate de răspuns, orice neclaritate va fi trasată în sensul imaginii respective, adică în direcția unor răspunsuri așteptate de la acest personaj social (anticipațiile de rol). În fine, dacă înaintea anchetei operatorul are o anumită ipoteză asupra distribuției răspunsurilor la unele întrebări (de exemplu va considera că trebuie să găsească un procent mare de nemulțumiți în legătură cu o problemă oarecare), atunci el va încerca, probabil, ca toate răspunsurile ambigue să le claseze de așa manieră încît să se încadreze în imaginea lui inițială. Un alt aspect, consider eu de aceeași natură, întîlnit frecvent în anchetele realizate cu studenți în calitate de operatori, apare cînd, după realizarea unui număr de chestionare, se constată că la anumite întrebări răspunsurile se dovedesc a fi uniforme, adică o variantă este aleasă de un procent foarte ridicat de indivizi. În asemenea situații, operatorul are tendința în continuare să claseze identic răspunsurile celorlalți subiecți. În ambele cazuri se poate vorbi de anticipație de probabilitate. Tot din experiența personală cu operatorii studenți pot mărturisi că uneori aceste anticipații, mai ales forma ultimă menționată, sînt atît de puternice încît operatorul nici nu se mai străduiește să adreseze unele întrebări din chestionar, completînd automat răspunsul. Probabil că în 99% din cazuri el nu greșește, dar niciodată nu sîntem siguri ce va spune



un om și deci o asemenea practică trebuie considerată, în principiu, ca generatoare de erori.

În legătură cu modalitatea de adresare a întrebărilor este greu de a detecta și clasifica riguros toate aspectele ce pot provoca erori. Uneori, la întrebările de opinie, este dificil de precizat dacă un anumit fel de a te adresa conduce la eroare sau mai degrabă la un răspuns corect. Așa cum am mai spus, răspunsul la aceste întrebări este foarte sensibil la modul de formulare și de aceea în multe anchete se cere operatorilor utilizarea unei forme standard, spre a se reduce măcar variabilitatea răspunsurilor datorată acestui fapt. Dar o astfel de formă nu este și nu poate fi niciodată respectată riguros de toți operatorii, esențialul rămânând ca abaterile individuale să se compenseze reciproc.

### 3.3. Erori datorate subiectului și contextului social în care se realizează ancheta

Probabil că dintre toate erorile ce afectează o cercetare, cele mai puține sînt cele ce pot fi puse precumpănitor în seama anchetatorului, căci, așa cum judicios remarcă Moser: "ca și în viața de toate zilele, ceea ce oamenii spun și cum spun depinde de împrejurări și de persoana cu care vorbesc" (1967, p.358). Am putea totuși, într-o primă instanță, clasifica erorile datorate subiectului în două categorii: conștiente și inconștiente, sau poate mai bine spus, intenționate și neintenționate.

Fiecare din aceste tipuri se regăsește atît în cazul întrebărilor factuale cît și în cel al întrebărilor de opinie. Dacă omul nu dorește să dezvăluie anumite stări de fapt (de exemplu, veniturile) sau anumite opinii el va da răspunsuri nu în conformitate cu situația reală, ci cu una ipotetică pe care o consideră a fi cea mai adecvată în contextul respectiv (va minimaliza veniturile, va declara opinii care să se încadreze în modelul dezirabilității sociale etc.) Erorile neintenționate pot fi datorate înțelegerii greșite a conținutului întrebării, faptului că nu-și aminteste corect anumite evenimente etc.

Erorile ce provin din răspunsurile neadecvate ale subiectului se datorează în bună măsură structurii psihice a acestuia. Memoria, atenția, voința, curajul, lipsa de prejudecăți etc. sînt



cîțiva din factorii de personalitate ce-și pun amprenta asupra comportamentului subiectului în situația de interviu. Nu este de neglijat nici influența contextului social în care se realizează ancheta. Anumite presiuni sociale de moment pot produce distorsiuni serioase în răspunsuri. De altfel, problema aceasta este ceva mai generală; anchetele sociologice sînt ele însele fapte sociale cu care oamenii sînt mai mult sau mai puțin familiarizați. De regulă, într-o societate în care nu există tradiția unor asemenea investigații, persoanele sînt surprinse și bănuitoare atunci cînd se trezesc în fața "omului cu chestionarul". Este nevoie deci de consolidarea unei experiențe sociale, de înțelegerea de către toată lumea că anchetele sociologice sînt inofensive față de un individ sau altul ca persoană privată, că ele slujesc în societatea noastră unor scopuri nobile, de cunoaștere și de ameliorare a vieții tuturor.

#### 3.4. Erori introduse în faza de prelucrare a datelor

Două ar fi sursele de erori în faza finală a unei anchete sociologice. Mai întîi, erorile pot apare din greselile săvîrșite în operația de postcodificare a întrebărilor deschise. Am amintit deja cît de multe dificultăți se ridică la prelucrarea acestor întrebări; oricît de riguroasă ar fi mulțimea categoriilor stabilite pentru încadrarea răspunsurilor respective, este practic imposibil să se asigure prelucrări o fidelitate foarte ridicată. Așa cum o demonstrează cu claritate rezultatele experimentelor efectuate în acest sens, repetarea codificării informației conținută în întrebările deschise, cu aceeași grilă de categorii dar de către alte persoane - sau chiar de o aceeași persoană după un timp oarecare - conduce la rezultate sensibil diferite (vezi Moser, 1967, pp. 388 - 393).

O altă operație efectuată în faza finală a cercetării, care este întotdeauna generatoare de erori, este cea de transpunere a informației de pe chestionar în instrumentele de prelucrare a datelor. Astfel, dacă chestionarul se prelucurează la calculator, toate codurile sale pot fi trecute inițial pe un formular standard după care se perforează cartelele; uneori, dacă chestionarul nu es-



te foarte complicat și mai ales dacă sistemul de codificare e uniform și clar, codurile pot fi trecute direct din chestionar pe cartele. Deși riscul de a greși, pentru persoanele ce perforază cartelele, este mai mare fiind lucrează direct cu chestionarul, eliminarea unei operații intermediare conduce cu siguranță și la eliminarea multor erori. De asemenea, și în cursul prelucrării manuale a datelor este posibil să se introducă erori; din fericire, aici ele pot fi descoperite relativ ușor și corijate așa precum, în cursul acestei prelucrări, mai pot fi detectate și corectate unele greșeli introduse în fazele anterioare ale cercetării.

Se poate aprecia că în faza prelucrării datelor - manuală sau automatizată -, exceptând desigur cazurile de neglijență crasă, erorile sînt relativ puține și nu pot avea o influență semnificativă asupra rezultatelor. Ba, mai mult, dacă dorim o precizie foarte mare, cu eforturi nu tocmai deosebite erorile pot fi aproape complet eliminate. De pildă, la preluorarea la calculator, cea mai simplă și eficace măsură de a elimina greșelile intervenite în transpunerea informației pe cartele constă în perforarea, direct de pe chestionar, a două rînduri de cartele - dacă e posibil de către operatoarele diferite - și compararea, tot cu ajutorul ordinatului, a perechilor de cartele corespunzînd aceluiași chestionar; se va pune imediat în evidență orice deosebire între două cartele paralele, identificîndu-se astfel erorile.

Este clar că o serie de erori ce intervin în faza de prelucrare a datelor își au izvorul în modul în care informația este înregistrată pe chestionar. Pentru a evita asemenea greșeli și chiar pentru a corecta unele erori introduse în faza de culegere a informației, înainte de preluorare, chestionarele trebuie supuse unei operații riguroase de verificare. Sub nici un motiv, chiar dacă se dispune de anchetatori foarte pricepuți, chestionarele nu vor fi trimise la prelucrare fără ca un grup, cît mai restrîns posibil, de cercetători să treacă la o verificare a lor. În cursul acestei



operații se urmăresc, în esență, trei lucruri de bază: completitudinea, exactitatea și uniformitatea înregistrării datelor. (Pentru detalii, a se vedea Moser, 1967, pp.379-383). Se vor elimina astfel erori însemnate datorate unor omisiuni în transcrierea informației, unor inadvertențe în răspunsurile la întrebările ce se controlează reciproc și a unor interpretări neuniforme a sensului anumitor întrebări. Majoritatea acestor erori, depistate înaintea prelucrării, pot fi eliminate pe cale logică sau prin reanchetarea unor subiecți.

Trecînd în evidență sursele erorilor în anchetele sociologice, se poate crea impresia că datorită numărului foarte mare al acestora și a intensității cu care unele acționează ar trebui să punem serios sub semnul întrebării eficiența metodei, validitatea rezultatelor obținute. Desigur, erorile în orice anchetă au o pondere deloc neglijabilă și nu e cazul nici să idealizăm metoda trecînd sub tăcere aspectele care nu ne plac. Totuși, acesta nu e un motiv suficient de puternic pentru a abandona una din metodele de bază ale sociologiei. Mai întîi, pentru că, dacă o cercetare este condusă cu seriozitate, erorile nu pot fi atît de mari încît să denatureze total sensul și semnificația unor constatări la care ea ajunge. Chiar dacă proporțiile sau mediile unor mărimi determinate prin anchetă nu sînt identice cu cele "reale" abaterile nu vor depăși o limită de toleranță acceptabilă. Găsind, de exemplu, că 75% dintre indivizii chestionați își exprimă insatisfacția față de un anumit aspect al vieții lor, înseamnă chiar că ne găsim în fața unei probleme sociale ce reclamă o rezolvare rapidă, chit că procentul real nu e 75, ci poate 70 sau 80. Importă mai puțin valoarea exactă a unui asemenea procent și mai mult ordinul său de mărime asupra căruia nu se poate chiar atît de ușor greși.

În al doilea rînd, realizarea anchetei pe o masă mare de indivizi face posibilă acțiunea legităților statistice, nu numai în scopul desprinderii unor tendințe în populație, dincolo de comportamentele individuale, dar și referitoare la erorile accidentale, nesistematice. Vreau să spun că, pe astfel de populații mari, o bună parte din erorile efective produse în fiecare chestionar, numite erori brute, se vor elimina prin anihilări reciproce, în calculul unor mărimi medii la nivelul lotului cercetat,



și vor rămâne erori finale mai reduse, numite erori nete.

În concluzie, se poate deci aprecia că problema erorilor în cercetarea de tip anchetă este una fundamentală, căreia sociologul trebuie să-i acorde întreaga atenție, de la începutul și pînă la sfîrșitul cercetării; nici un efort în acest sens nu va rămîne nerăsplătit. Într-o asemenea strategie generală locul central trebuie să-l ocupe depistarea factorilor ce produc erori sistematice, a căror eliminare nu este de așteptat din partea hartzului; ei sînt legați, așa cum s-a văzut, în special de modul de întocmire a chestionarului și de munca operatorilor.



## Cap. V. ANALIZA DOCUMENTELOR

O altă metodă specifică sociologiei - și în general științelor socio-umane, - alături de anchetă, o constituie tot o specie derivată din metoda generală a observației, aplicată însă nu la realitatea socială în sine, ci la o imagine a ei, și anume la documentele produse și reproduse în fiecare tip de societate umană. Vom numi metoda cu termenii din titlu: analiza documentelor.

Lecția de față urmărește să prezinte trei probleme fundamentale: ce, unde și cum se investighează cu metoda în cauză, adică să specifice conținutul documentelor care constituie obiectul analizei sociologice, tipul de documente în care acest conținut este identificabil și tehnicile de analiză subordonate metodei. Primele două aspecte vor fi schițate împreună, dat fiind că despre conținutul documentelor cu greu s-ar putea vorbi în abstract, fără a face referiri la natura acestora, din cauza enormei varietăți de informații documentare ce sînt utile sociologului, în nenumărate probleme studiate de el.

Cît privește rostul și importanța metodei în științele sociale, nu cred că este nevoie să se spună prea multe; oricine înțelege ușor că informația documentară este una absolut indispensabilă "observatorului" faptelor sociale, oă neglijarea ei și reducerea arsenalului metodologic al sociologiei doar la observații directe sau anchete înseamnă renunțarea cu bună știință la o serie de date și informații ce nu pot fi reconstituite cu alte mijloace. De altfel, se poate aprecia oă analiza documentelor a constituit prima metodă majoră de cercetare în sociologia nespeculativă. Lucrări clasice în domeniu precum Capitalul lui Marx, Situația clasei muncitoare din Anglia a lui Engels, Le Suicide de Durkheim sau scrierile lui Lenin despre dezvoltarea capitalismului în Rusia, ca să pomenesc doar cîteva exemple mai semnificative, sînt în întregime elaborate în substanța lor informațională cu ajutorul metodei prezentată.

Odată cu dezvoltarea sociologiei empirice de teren, alte



metode au trecut în prim-planul atenției sociologilor: ancheta, interviul sau observația. Cu toate acestea, se poate afirma că analiza documentară n-a fost niciodată abandonată chiar dacă a fost folosită numai ca însoțitoare a altor metode și asta cel puțin pentru un motiv esențial: înțelegerea fenomenelor prezente nu este, de regulă, satisfăcătoare fără a cunoaște evoluția ori geneza lor, deci trecutul, formele de manifestare pînă în momentul cercetării. Or, despre trecut, metoda care furnizează datele cele mai consistente este tocmai analiza documentară.

După părerea mea, la ora actuală, cuplul metodologic anchetă-analiza documentelor constituie instrumentul cel mai eficient de cunoaștere pentru sociologie. Desigur, sînt situații în care numai una dintre cele două metode poate fi utilizată, dar nu este mai puțin adevărat că deseori ancheta este pornită pe un teren gol, nesustînută de alte surse de cunoaștere, și datele furnizate de ea sînt inconsistente, ajungîndu-se astfel la discreditarea acestei importante metode de cunoaștere a vieții sociale.

Legitimitatea analizei documentelor, ca metodă sociologică, se află, în ultima instanță, în faptul că documentele - indiferent de natura lor - sînt produsul activității umane, produs în care se încorporează o anumită cantitate de informație despre faptele și fenomenele sociale din perioada în care ele au fost produse. La prima vedere poate părea că metoda ar fi proprie doar istoriei și nu sociologiei căci documentele dau seama totdeauna despre un trecut. Argumentele împotriva acestei idei sînt însă imediate:

(i) Sociologia, în accepția multora dintre cei care o practică, nu este numai știința prezentului social, ci și a trecutului, fiind complementară studiului istoric al societăților pe axa cronologică.

(ii) În al doilea rînd, așa cum aminteam deja, cunoașterea trecutului este deseori indispensabilă pentru înțelegerea corectă a prezentului.

(iii) Chiar dacă nu acceptăm ideile de mai sus în toată amploarea lor, va trebui totuși admis că prezentul nu se reduce la o clipă ce trece rapid, ci are și el o durată, deci un trecut mai lung sau mai scurt ce-i aparține în fapt. Sau, altfel spus, docu-



mentele nu sînt caracteristice numai trecutului, ci și prezentului; ele dau seama, uneori foarte bine, despre ceea ce se întîmplă acum și aici.

(iv) În fine, metoda documentară nu ține doar de studiul istoric chiar și pentru simplul motiv că ea se aplică, așa cum vom vedea imediat, chiar la "obiecte" create de cercetarea sociologică (documente constituite pentru a fi studiate sociologic).

Dar mai bine să începem prezentarea ariei de aplicare a metodei, căci atunci și rolul și importanța pentru sociologie, dar și specificul său în cadrul disciplinei noastre se vor lămurii, sper, pe deplin.

### 1. Tipuri de documente

Fără prezența de a da o definiție riguroasă, vom înțelege prin document orice obiect material realizat de om, capabil să imaginezeze o anumită informație. O precizare se impune însă imediat. Practic orice lucru construit sau chiar atins de mîna omului, ca să spun așa, va reține o informație despre cel care l-a realizat sau manipulat. O unealtă de muncă, o clădire, un drum, o simfonie, o armă, un poem etc. sînt cîteva exemple de entități create de om ce pot fi - și sînt efectiv - folosite în reconstituirea vieții sociale din epoca respectivă. Dar nu toate aceste documente au aceeași relevanță pentru sociolog ca și pentru istoric sau etnolog, chiar dacă nu este posibil să se traseze aprioric o linie despărțitoare în afara căreia să plasăm exact toate obiectele ce vor intra în sfera de preocupări a sociologului, cu titlul de documente. În principiu deci vom admite că orice astfel de entitate poate deveni un document sociologic, dar vom înțelege că, în realitate, sfera celor care sînt efectiv utilizate aici e mult mai redusă, sau că există o anumită ierarhie, destul de greu, de altfel, de precizat și ea, a utilității documentelor pentru sociolog.

În cele ce urmează voi da o clasificare a documentelor, apropiată de cea propusă de alți autori, precum M. Duverger (1964) sau M. Grawitz (1972), rostul ei fiind doar acela de a introduce o anumită ordine sau de a constitui un fir conducător pentru o trecere în revistă cît mai completă a tipurilor de documente utilizabile.



### 1.1. Documente scrise de natură nesociologică

a) Statisticile oficiale. Sînt documente, cu caracter oficial, întocmite de unele organisme, pe baza numărării elementelor ce prezintă anumite caracteristici. Exemplul tipic de astfel de documente îl reprezintă rezultatele recensămintelor. Aceste operații se efectuează, la ora actuale, aproximativ din 10 în 10 ani, în mai toate țările lumii, avînd drept scop furnizarea unor informații demo-sociale necesare pentru factorii de conducere în stat. În intervalele dintre recensăminte unele date - ca cele referitoare la mișcarea populației, de exemplu - sînt reactualizate, an de an, pe baza înregistrărilor statistice curente. În cazul datelor demografice de bază, informațiile pot fi ținute la zi; cunoscînd structurile populației la recensăminte și evenimentele demografice din cursul unui an, se pot calcula relativ ușor volumul și structura populației în orice moment dorim.

La noi în țară, statisticile curente anuale se publică în Anuarul statistic, care cuprinde, în cea mai mare parte, indicatori de factură economică, dar și unele date demografice și geografice de bază. Detaliile de prezentare coboară pînă la nivelul județelor și/sau mediului urban și rural. Pe lângă lucrarea pomenită, Direcția Centrală de Statistică mai publică, din cînd în cînd, culegeri de date statistice pe anumite domenii, editează Revista de statistică, unde pot fi găsite, de asemenea, informații statistice oficiale.

Importanța surselor statistice pentru sociolog este extrem de mare, deoarece de aici el obține date relativ sigure și cu o arie de cuprindere maximă asupra societății în cauză. Informația aceasta primară, supusă unor prelucrări statistice adecvate și culeasă pe o secvență temporală suficient de lungă, poate constitui un obiect perfect de analiză sociologică în sine sau poate folosi la conturarea unui cadru de referință necesar în evaluarea datelor culese prin alte metode. Din păcate însă, aceste calități sînt de multe ori contrabalansate de penuria informațiilor oficiale care să intereseze în mod deosebit pe sociolog; așa cum spuneam, statisticile oficiale conțin - și la noi și aiurea - foarte multe informații de natură îngust economică și foarte puține referitoare la alte aspecte sociale.



b) Arhive publice și alte documente oficiale. Arhivele sînt colecții de documente al căror conținut poate ajuta sociologul să înțeleagă o serie de fenomene ale vieții sociale; ele sînt evident mai folositoare atunci cînd dorim să cercetăm evoluția sau geneza unor fenomene, decît cînd abordăm trecutul unor comunități umane. Este clar că arhivele sînt mult mai des frecventate de către istorici decît de sociologi, dar nu sînt ocolite nici de aceștia din urmă.

Un interes foarte mare pentru sociolog îl prezintă documentele oficiale ale statului ce sînt aduse la cunoștința publicului prin Buletinul Oficial, și, cele mai importante, prin mijloacele de comunicare de masă. De o deosebită importanță sînt la noi în țară documentele Partidului Comunist Român, elaborate cu ocazia plenarelor Comitetului Central, a conferințelor naționale ale partidului, a congreselor. În aceste documente apar, pe lîngă date privind evoluția unor indicatori socio-economici, și analize aprofundate ale dezvoltării societății noastre, deci un întreg arsenal de concepte și idei privind conducerea, organizarea și evoluția vieții sociale în România.

c) Presa. Prin caracterul său de oglindă a faptelor cotidiene, de mai mică sau mai mare importanță socială, presa este, în majoritatea țărilor, un izvor de informații de o bogăție inepuizabilă pentru cel care dorește să descifreze prin intermediul ei anumite probleme ale unei comunități. Desigur, există diferențe sensibile de la țară la țară în privința modului în care presa oglindește cu fidelitate faptele, dar oricîte obstacole ar ridica orientarea politică sau ideologică a diverselor ziare, pînă la urmă o analiză atentă poate să descrieze esențialul.

Caracteristic pentru presa din țara noastră este faptul că ea reflectă, în cea mai mare parte a conținutului său, aspecte majore ale societății românești de astăzi, putînd deci conduce sociologul spre cunoașterea unor elemente de mare profunzime. Dar chiar realizată pe laturile mai puțin importante, cum ar fi anunțurile la mica publicitate, faptele diverse etc. analiza de conținut a ziarelor nu este lipsită de interes.

d) Documentele instituțiilor, organizațiilor, întreprinderilor etc. dau, în marea lor majoritate, informații despre modul de organizare, de funcționare, despre numărul și statutul persoanelor



ce iau parte la activitățile respective, despre relațiile cu alte unități și, în general, despre locul și rolul jucat de acestea în viața socială. Așa de exemplu, analiza proceselor verbale din ședințele C.O.M.-ului unei întreprinderi poate spune lucruri foarte interesante sociologului, chiar făcând abstracție de conținutul tehnic al discuțiilor purtate aici - și mai ales făcând abstracție de acesta. Se relevă astfel aspecte semnificative ale procesului de luare a deciziilor, gradul de participare la acest proces a reprezentanților muncitorilor, stilul de muncă al conducătorului întreprinderii etc. Nu mai vorbesc de faptul că documentele aflate la serviciul personal al unității sînt surse fără a căror consultare sociologul industrial nu poate practic studia o serie de probleme cum ar fi fluctuația forței de muncă și altele.

Exemplul cu întreprinderea, deși foarte sugestiv, este departe de a se încadra în rîndul celor mai semnificative cazuri, deoarece documentele ce se găsesc aici vizează un sector de activitate relativ restrîns și un volum de persoane modest ca mărime. Alte unități posedă documente cu deschidere mult mai amplă; mă gîndesc la valoroasa (pentru sociolog) documentație ce se găsește la diverse organe ca cele care dirijează, pe raza unui județ, de exemplu, activitatea școlară, sanitară, financiară, judecătorească etc. Cu siguranță că informațiile respective sînt folosite de organele de specialitate, dar sociologii le-ar putea valorifica din alte puncte de vedere, găsindu-le sensuri mai largi, descoperind corelații, determinări sau dependențe între fenomene ce ies în afara unei ramuri înguste de activitate. Din păcate, astfel de surse sînt rareori exploatate în cercetarea sociologică - și nu numai la noi - lucru care se datorează în principal faptului că accesul la astfel de documente nu este ușor.

e) Documente personale. Este vorba aici de scrisori, jurnale, autobiografii, certificate, diplome etc., deci de tot ceea ce poate da seama despre viața unei persoane sau a familiei sale, despre raporturile între indivizii sau microgrupuri sociale, dar și de materiale ce conțin relatări despre fenomene cu o sferă mai largă, cu implicații sociale mai profunde.

Referitor la această sursă informațională, toate manualele



de metode pomenesc ca exemplu, ca model de analiză deosebit de reușită celebra lucrare a sociologilor Thomas și Znaniecki, Țăranul polonez, apărută în 1919 în S.U.A. Studiind un număr impresionant de documente personale (scrisori, jurnale, autobiografii etc.), dar și de altă natură (articole de ziare, arhive parohiale, rapoarte ale tribunalelor sau ale organismelor de ajutor social), autorii au urmărit, în special, să pună în evidență procesul de transformare a satului polonez ca urmare a schimbărilor survenite, la începutul secolului, în sistemul de valori și atitudini al țăranilor, schimbări accelerate și de migrația masivă în Statele Unite. Cei doi sociologi consideră că documentele personale, în speță autobiografiile, constituie tipul perfect de material sociologic.

f) Literatura poate fi considerată fără nici o rezervă drept sursă de informații asupra vieții sociale dintr-o anumită epocă, indiferent de interpretările și concepțiile personale ale autorilor. Pentru a ne da seama de forța pe care literatura o posedă în această privință este suficient să ne gândim - desigur cei care nu sîntem istorici de meserie - la societatea rusă sau la cea franceză din secolul trecut și, încercînd să decelăm sursele cunoștințelor noastre despre ele, vom ajunge la concluzia că sîntem în proporție covîrșitoare tributari imaginilor pe care marii romancieri ai celor două literaturi ni le-au transmis.

g) Alte lucrări tipărite precum dicționarele, enciclopediile, afișele etc. sînt considerate, de asemenea, ca obiecte purtătoare de informație relevantă pentru sociolog.

## 1.2. Documente scrise sociologice

Voi cuprinde sub acest titlu toate documentele scrise elaborate în scop sociologic, în cadrul unor cercetări empirice. Ele pot fi de două feluri:

- materiale ce provin din cercetări anterioare celei în curs, cercetări efectuate fie de sociologul în cauză, fie de către alții, și care se găsesc sub formă de rapoarte de cercetare, articole cărți, arhive sociologice etc. tratînd o problemă identică sau apropiată de cea urmărită;



- materiale constituite în cursul investigației respective, ce se cer ele însele cercetate cu tehnici ale analizei documentelor ( chestionare cu întrebări deschise, interviuri, etc.)

Cîteva precizări sînt necesare aici. De regulă, manualele nu includ asemenea documente în clasificările lor, fie probabil din cauza unor omisiuni involuntare, fie poate datorită temerilor de a nu extinde prea mult aria metodei, ceea ce ar face, la urma urmei, să-și piardă specificitatea și să poată fi invocată în orice gen de cercetare. Mai întîi, deci, trebuie să înlăturăm tocmai o asemenea posibilitate. Este clar, cum spuneam și cu altă ocazie, că studiul bibliografiei este un punct important în pregătirea oricărei cercetări și, prin urmare, dacă analiza documentelor rămîne o metodă specifică în cercetarea sociologică, ea nu trebuie să se suprapună cu acest studiu. E vorba în cazul metodei aici prezentate de o utilizare a informațiilor empirice din alte cercetări sociologice, la fel cum astăzi de informații sînt căutate în orice document, în timp ce prin studiul bibliografic urmărim cu precădere aspecte teoretice (precizarea noțiunilor, formularea ipotezelor etc.) și doar parțial date empirice, și acestea numai cu scopul de a găsi elemente de comparație pentru datele cercetării noastre iar nu ca fapte spuse la rîndul lor investigației.

Reamintin că obiectul cursului acesta îl constituie prezentarea modalităților de cercetare sociologică empirică, vom observa totuși că în cazul aplicării analizei documentelor la materiale sociologice, distincția empiric-teoretic nu este foarte netă. Orice studiu sociologic pe care-l vom folosi ca sursă informațională este, într-o proporție sau în alta, un amestec de informații empirice și constructe teoretice care au ghidat cercetarea în cauză. Ar fi absurd să se creadă că cele două elemente sînt detașabile și că s-ar putea extrage date empirice în formă pură. Dar aceasta nu ne împiedică totuși să distingem cele două unghiuri principale din care sînt privite documentele sociologice: ca surse de informații despre realitatea studiată și ca teorie sau metodologie de studiere a aceleiași realități.

Așadar analiza secundară a datelor sociologice poate fi considerată în mod cu totul îndreptățit ca o metodă (variantă a metodei analizei documentelor) importantă în arsenalul cercetării



sociologice. Mai mult, această metodă nu este destinată doar să acompanieze alte metode într-o investigație oarecare, ci poate constitui singură o cale eficientă de a obține rezultate relevante și importante, derivate din cercetări sociologice anterioare, dar nicidecum reductibile la concluziile acestor cercetări. Este limpede că dacă sociologul dispune de o masă cât mai mare de informații empirice (comparabile) despre un anumit fenomen, el va avea șanse sporite de a decela unele regularități, conexiuni etc. care n-au putut fi puse în evidență de nici un material sociologic care i-a stat la bază. Un exemplu elocvent în această privință îl constituie lucrarea lui R. Boudon L'Inégalité des chances (1973), unde sociologul francez sintetizează rezultatele a zeci de cercetări asupra inegalităților de șanse în fața școlii și creionează implicațiile acestora asupra mobilității sociale.

Referitor la al doilea punct menționat (analiza documentelor produse pe parcursul cercetării în cauză), se vede că aici metoda este utilizată doar în prelungirea altei modalități de culegere a datelor (ancheta sau interviul) și nu în mod independent. Textele lungi ale unor întrebări deschise din chestionar sau ale interviurilor apar ca documente de sine stătătoare asupra cărora sociologul trebuie să se aplece cu instrumente de cercetare adecvate.

### 1.3. Alte surse

În afara documentelor scrise, există și alte documente ce intră în câmpul analizei sociologice; e drept însă că folosirea lor este incomparabil mai redusă în disciplina noastră decât în alte științe sociale învecinate: istoria, etnografia, antropologia etc.

Asemenea surse le-am putea clasifica, împreună cu Duverger (1964), în felul următor:

- a) Obiecte tehnice (unelte, clădiri, mobilier etc.)
- b) Documente iconografice (desene, gravuri, tablouri, picturi), fotografice și cinematografice.
- c) Documente fonetice (discuri, benzi de magnetofon, casete).



În exemplificările de până aici am sugerat tot timpul o singură perspectivă din care se efectuează analiza documentelor, și anume acea care vizează conținutul documentelor ca reflectare, reproducere, înregistrare, adică drept sursă de informații despre anumite fapte sociale exterioare și anterioare documentelor respective.

Se impune însă cu necesitate precizarea unui al doilea punct de vedere din care documentele îl interesează pe sociolog, aspect ce, în multe situații, poate să fie chiar mai important decât cel precedent. Se înțelege ușor că orice document este el însuși un "obiect" social, la elaborarea cărui a participat un individ, un grup, o instituție etc. Deci documentul este rezultatul unei activități umane, determinată de anumite fapte, împrejurări sociale, mobilizând o serie de forțe omenesti în realizarea lui și care, odată realizat, mai cu seamă dacă posedă o largă circulație, va fi generator de efecte sociale.

Așadar, o a doua perspectivă sociologică va fi cea care încearcă să dea răspunsuri la întrebări precum: cine a produs documentele, cu ce scop, cui i se adresează informația conținută, ce gen de informație manipulează unele documente în comparație cu altele, ce efect social a produs elaborarea și circulația unui document etc. În consecință, metoda analizei documentare cunoaște două variante de bază: una pe care o voi numi, în lipsă de un termen probabil mai adecvat, directă, utilizează în mod nemijlocit informația conținută în documente ca atare, deci ca informație despre fapte sociale ce nu au altă legătură cu documentele decât că acestea din urmă dau seama despre primele și alta, să-i zicem indirectă, în care documentul însuși are o semnificație socială, iar informația conținută în el cere o interpretare deosebită, mediată de contextul social în care documentul a fost elaborat.

Practic, fiecare gen de documente este capabil să ne furnizeze informații indirecte; unele precum obiectele tehnice, sînt aproape în exclusivitate purtătoarele unei astfel de informație. Dar chiar și cele mai importante documente, din punctul de vedere al bogăției de cunoștințe directe despre societate, se pretează la fructuoase analize sociologice din unghiul al doilea menționat. Astfel, presa în țările occidentale constituie un element privilegiat de analiză documentară indirectă. Studiind, de exemplu, relațiile pe care diferite jurnale le dau aceluiași eveniment, se pot trage concluzii interesante despre orientarea ziarelor în cau-



ză, tot așa cum analiza conținutului unei publicații (genul rubricilor sale, extinderea acestora, natura valorilor promovate ș.a.m.d) este în măsură să ofere cunoștințe asupra publicului cărui i se adresează, asupra ideologiei promovate, deci asupra intereselor de clasă sau grupele pe care le apără. Cercetările asupra conținutului filmelor scot în evidență, de asemenea, o mulțime de lucruri importante, fie pe planul evoluției unor valori sau norme sociale, al tipului de eroi ori de conflicte promovate în cadrul unei cinematografii naționale, fie diferențele - după parametri similari - între producțiile cinematografice din două țări la un moment dat. Natural că și programele posturilor de radio sau televiziune se pretează din plin la acest gen de analize.

Un domeniu în care analiza de tip indirect a documentelor își găsește un loc de seamă îl constituie sociologia politică. Discursurilor liderilor politici, de pildă, sînt studiate extrem de minuțios: apariția unor teme sau cuvinte cheie noi ori repetarea unora mai vechi sînt indicatori principali după care se poate judeca orientarea politică a conducerii partidelor. Programele, platformele electorale, afișele sau lozincile partidelor nu conțin numai informații sociale generale - și poate nu în primul rînd astfel de informații -, ci indicii asupra orientării, a poziției sociale a partidului respectiv.

Chiar și lucrările de sociologie, atît empirice cît și teoretice - ca, de altfel, orice alt gen de lucrări științifice - constituie obiect de analiză indirectă, fiind astfel purtătoare de informații pentru sociologia sociologiei (sau sociologia științei, mai general). Analiza conținutului revistelor sociologice, după tematica abordată ori după metodele utilizate în cercetare, studierea întregului volum de publicații sociologice într-un interval de timp și într-o țară anume, analiza trimitelor bibliografice dontr-un set de lucrări, după naționalitatea autorilor sau limba în care au scris, după anul apariției etc. sînt doar situațiile cele mai frecvent întîlnite ce se încadrează perfect în problema aici descrisă.

Asemenea ilustrări ar putea continua pe mai multe pagini, dar exemplele precedente cred că sînt concludente pentru a reflec-



ta bogăția de perspective din care documentele pot fi analizate sociologic, indiferent de natura sau destinația lor. Probabil că dihotomia mai sus menționată (analiză documentară directă și indirectă), ca orice dihotomie, este un pic rigidă și ar fi nevoie de eforturi substanțiale de clasificare a tuturor unghiurilor de vedere, a tuturor pistelor ce-i sînt deschise sociologului în elaborarea documentelor ca obiect de analiză. Deși idei similare pot fi găsite în multe cursuri și tratate, în sensul că anumite genuri de documente se pretează la mai multe posibilități de analiză, nicăieri, personal, nu am găsit o abordare generală a analizei documentare după asemenea posibilități. Iar distincția operată aici între o analiză directă (care merge pe linia folosirii informației din documente ca atare, deci ca informație despre anumite fapte sociale petrecute) și una indirectă (în care documentul este un fapt social sau conținutul să nu interesează atît ca informație asupra evenimentului relatat ei, în primul rînd, asupra celui care relatează, a felului în care relatează, vizînd, prin urmare, o informație dincolo de cea aparentă) această distincție, zic, chiar dacă poate prea abruptă, mi se pare totuși esențială.

De aici decurge și o problemă metodologică importantă. În primul tip de analiză, cercetătorul trebuie să aibă mereu în vedere și să verifice dacă informația conținută în materialele ce urmează a fi investigate este corectă; el va exclude din capul locului din câmpul de cercetare orice document cu un conținut îndoielnic iar atunci cînd bănuiește existența unor distorsiuni, în lipsă de alte surse mai adecvate, va încerca să le corecteze. Astfel, dacă - așa cum se întîmplă frecvent în țările occidentale - mai multe ziare dau relatări și interpretări diferite aceluiași eveniment, prin compararea unei game largi de texte și eliminarea distorsiunilor datorate orientării ziarelor, sociologul poate ajunge uneori la o informație relativ obiectivă.

În al doilea gen de analiză, asemenea probleme nu se pun deci conținutul documentelor nu va fi judecat în termeni de adevărat sau fals, ceea ce face ca sfera de cuprindere să fie aici potențial mai largă decît în cazul precedent. Singura problemă ce ține de alegerea documentelor este aceea de a alege surse capabile să ofere un suport solid de cercetare, adecvat scopului ce și l-a propus sociologul.



## 2. Tehnici de analiză

După rapida trecere în revistă a "obiectelor" de analiză și a punctelor de vedere din care ele sînt interesante și importante pentru investigația sociologică empirică, urmează în mod logic să vedem cum se realizează concret analiza, care sînt tehnicile și instrumentele atașate metodei. Din păcate un studiu sistematic în acest sens este greu de întreprins, căci puține sînt elementele specifice muncii sociologului în procesul de studiere a documentelor; vreau să spun că activitatea lui în acest domeniu nu se deosebește din punct de vedere metodic de cea a istoricului a criticului literar sau a cercetătorilor din alte domenii.

Desigur, vorbind despre modalitățile de fructificare a informației din documente, trebuie să detașăm, în primul rînd, acele surse în care informația sociologică apare sub formă numerică, de regulă rezultat al numărării și, mai rar, al măsurării unor caracteristici. Astfel de date intră imediat și firesc în angrenajul clasic de analiză statistică, asupra căruia ne vom opri în partea a treia a cursului nostru. Voi menționa aici doar faptul că sociologul trebuie să se manifeste extrem de prudent atunci cînd folosește date statistice culese de alte persoane. În principiu, utilizarea acestor cifre trebuie să fie precedată de cunoașterea exactă a modului în care ele au fost constituite, adică a unităților statistice numărate și/sau a însușirilor măsurate, precum și a gradului de precizie sau temeinicie cu care s-au realizat operațiile de mai sus. Nu puține sînt cazurile cînd, pornind de la cifre incorect constituite, se aplică proceduri statistice sau matematice foarte sofisticate ce nu pot să producă decît rezultate iluzorii. De asemenea, lipsa de unitate în definirea unităților de analiză face inutilizabile datele statistice de care sociologul dispune sau, și mai rău, în cazul cînd asemenea inadvertențe tres neobservate, comparațiile și concluziile vor fi, forțamente, false. A compara, de exemplu, nivelele de urbanizare în două țări, fără a ține seama de criteriile diferite după care o localitate primește statutul de oraș (se poate întîmpla ca în una să fie considerată urbană orice localitate ce depășește 2000 de locuitori și în alta pragul minim să fie urcat la 10.000), este evident un nonsens, oricît de precise ar fi statisticile ce ne stau la dispoziție. Sau, un



alt exemplu în legătură cu precizia datelor. S-a observat că în unele țări în curs de dezvoltare ratele de natalitate au manifestat o tendință de creștere în ultimele decenii, opusă deci celei așteptate. La o analiză atentă s-a dovedit că o serie de astfel de cifre ce datorează pur și simplu unei înregistrări mai corecte, pe măsura trecerii timpului, a cazurilor de naștere în aceste țări.

Luind la o parte documentele a căror analiză este transferabilă direct procedurilor statistice, vom observa că celelalte surse documentare pot fi supuse, în principiu, unor tehnici de analiză calitativă sau cantitativă. Tehnicile calitative sînt, evident, cele mai curent întrebuintate și în sociologie ca și în alte discipline socio-umane. Aici extragerea informației din documente nu face apel la instrumente deosebite, ci se folosesc cele clasice și banale: fișa, conspectul, rezumatul etc. Iar analiza urmează, de asemenea, căile cunoscute: remarcarea apariției unor informații, comparații privind conținutul documentelor diverse, detagarea ideilor principale într-un text, reconstituirea contextului social-istoric în care s-a elaborat documentul etc., fiecare cercetător (sociolog, istoric sau critic literar) efectuînd-o fără să-și conștientizeze în mod deosebit actele, așa cum - după remarcă lui M. Duverger - domnul Jourdain făcea proză fără, să știe.

Este normal că, în lipsa unor cadre obiective de analiză, oricît de conștient, informat sau bine intenționat ar fi un cercetător, rezultatele muncii sale nu sînt scutite de influențe subiective, de interpretări ale textelor determinate de condiția socială, morală sau intelectuală a celui care întreprinde analiza. Acest lucru este resimțit în toate domeniile care folosesc analiza documentelor, iar ultima jumătate de secol este caracterizată de eforturi multiple în depistarea acelor tehnici care să permită, în măsura posibilului, depășirea impresionismului (deci a subiectivismului) în astfel de analize. Una dintre direcțiile pe care s-a mers - și, repet, nu numai în sociologie - este cea a analizei de tip cantitativ, al cărei obiectiv fundamental îl constituie transformarea informației descriptiv-discursive a documentelor în unități (entități) numărabile, sau, eventual, măsurabile, după anumite însușiri. Tehnica impusă și perfecționată de sociologie și psihologie socială capabilă să satisfacă un asemenea deziderat poartă denumirea de analiză de conținut, iar obiectul rîndurilor



ce urmează îl constituie schițarea liniilor generale de punere în lucru a acestei tehnici.

Mai întâi însă iarăși o remarcă. În contextul actual, caracterizat de un contraatac destul de viguros la adresa pozitivismului din partea unor orientări filosofice opuse, și în sociologie, în cadrul mai larg al frontului anticantitativist, se pot întâlni frecvent idei care neagă orice valoare studiilor de tip cantitativ în analiza documentară, făcând apologia tehnicilor comprehensive, hermeneuticii etc. Este clar că și aici, ca și referitor la multe alte probleme, astfel de reacții sînt de înțeles iar critica apare justificată atunci cînd vizează excesele cantitativiste; dar nu o dată criticile sînt generale, fără discernămint și de aceea nocive. De asemenea, este iarăși limpede că o analiză strict cantitativă a documentelor este practic imposibilă și că în cadrul analizei de conținut, de pildă, aspectele cantitative și cele calitative se împletesc poate mai strîns decît în orice altă metodă sau tehnică de cercetare sociologică.

Revenind la analiza de conținut, vom accepta, pentru moment, în loc de definiție, următoarea caracterizare: ea este o tehnică de cercetare a documentelor ce furnizează o descriere obiectivă sistematică și cantitativă a conținutului acestora, urmînd ca aspectele concrete ale tehnicii să fie înțelese din descrierea ce urmează.

Nu voi intra în detalii privind istoricul problemei. Deși unele încercări de tratare cantitativă a textelor sînt cunoscute de multă vreme, lucrările de metodologie înclină, în majoritate, să-l considere drept părinte adevărat al analizei de conținut pe Harold D. Lasswell, un psihosociolog american, care a elaborat principiile analizei și a utilizat-o împreună cu colaboratorii, în domeniul sociologiei politice, în perioada celui de al doilea război mondial. Pornită deci inițial de la studiul comunicațiilor politice (discursuri, comunicate oficiale, știri din presa scrisă și vorbită etc.), analiza de conținut s-a dovedit a fi o tehnică ce se poate aplica și în alte domenii cu aceeași eficiență.

În principal aria de utilizare a tehnicii o reprezintă documentele scrise, iar aspectul cantitativ despre care aminteam este cel statistic (de numărare). Din acest motiv, pentru a fi posibilă aplicarea tehnicii este nevoie ca "populația" documentelor



ce intră în studiu să fie suficient de largă, altfel nu se pot pune în evidență regularitățile statistice. În cazul în care volumul documentelor ajunge să depășească posibilitățile unei analize exhaustive, se pot aplica și aici procedee de eșantionare, alegându-se un eșantion de texte care să posede o reprezentativitate corespunzătoare necesităților cercetării. De exemplu, dacă lotul integral de documente este reprezentat de numerele unui ziar, cotidian, extinse pe durata unui deceniu, vor trebui cercetate peste 3.100 de numere (presupunând că o zi pe săptămână ziarul nu apare), ceea ce, exceptând cazul unei analize foarte superficiale, e greu de realizat. Un eșantion de cca 500 de ziare (unul pe săptămână) poate fi suficient pentru atingerea scopurilor urmărite.

Stabilirea populației și a eșantionului de documente se face în corelație strânsă cu precizarea foarte exactă a problemei de studiat. În mod deosebit, în acest gen de analiză problema cercetată trebuie să primească o delimitare riguroasă și să fi transpusă corect și complet în indicatori empirici. Chestiunea centrală a studiului se concretizează, de regulă, în una sau mai multe caracteristici, a căror prefigurare cantitativă se urmărește în studiile mai simple (explorative); acestora li se adaugă și alte caracteristici, auxiliare, cu rol de variabile explicative, în studiile mai complexe.

De exemplu, un studiu explorativ poate avea drept scop evidențierea "tematicii articolelor de fond din ziarul Scînteia" pe o perioadă dată. Expresia dintre ghilimele este atunci singura caracteristică ce urmează a fi construită în urma studiului empiric. La fel și în analize cu teme ca: "valorile morale promovate de emisiunile de tineret ale televiziunii", "tipurile socio-profesionale de personaje din reportajele Scînteii tineretului", "ocupăția eroilor principali din romanul românesc din ultimii 35 de ani" etc.

Adîncirea analizei poate merge cel puțin pe două coordonate principale: fie o aceeași caracteristică se va studia pe subpopulații distinse de documente (de pildă, prima temă menționată va fi urmărită pe anumite perioade de apariție a ziarului Scînteia ori caracteristica va fi elaborată pentru ziare diferite etc.), fie populația rămîne aceeași și se construiesc mai multe caracte-



risticii pe ea, urmînd ca acestea să fie apoi studiate prin proceduri statistice precum asocierea sau corelația. A doua variantă se apropie de cercetările pe bază de anchetă, numai că dacă printr-un chestionar putem ajunge să obținem zeci sau chiar sute de caracteristici ale populației, în analiza de conținut cu greu se vor obține mai mult de 3-4 asemenea indicatori, din cauză că, pe de o parte, documentele nu sînt capabile să "răspundă" la atîtea întrebări ca oamenii iar, pe de altă parte, se ridică serioase dificultăți practice în construcția simultană a unui număr ridicat de caracteristici pe același documente.

Se înțelege ușor că mulțimea, mai largă sau mai restrînsă, a caracteristicilor propuse spre construcție este dependentă de conținutul informațional al documentelor, deci însăși problema de cercetat trebuie corelată corect cu sursa informațională. De asemenea, populația documentelor se cere circumscriasă cel puțin cu tot atîta atenție ca și populația umană în cadrul anchetelor. În fine, atunci cînd este vorba de documente ce apar într-o succesiune regulată, cum sînt de pildă ziarele, trebuie avut în vedere ca perioada analizată să fie semnificativă pentru faptele sociale urmărite. Spre exemplu, dacă ne va interesa felul în care Scînteia tineretului reflectă participarea tinerilor la campaniile agricole de recoltare, atunci se cer alese în fiecare an numere din perioadele cînd se realizează aceste campanii. Desigur, cazul acesta este banal, dar el reflectă o situație mai generală, și anume aceea că din cauza periodicității unor acțiuni sociale, trebuie acordată toată atenția felului în care se aleg documentele.

În continuare să presupunem, pentru simplificarea expunerii, că analiza de conținut ce urmează să o întreprindem vizează stabilirea unei singure caracteristici. După delimitarea exactă a populației de documente și precizarea semnificației caracteristicii în cauză, urmează operația de stabilire a categoriilor caracteristicii.

Categoriile sau clasele (voi folosi mai ales primul termen, încetățenit în literatura despre analiza de conținut) nu sînt altceva decît rubricile în care conținutul informației documentare este clasat pe baza unui criteriu. Altfel spus, categoriile sînt omoloagele variantelor de răspuns la o întrebare din chestionar, numai că de regulă, problemele cercetate prin analiza de conținut



sînt de asemenea natură încît avantajul variantelor este foarte complex. Din această cauză, se consideră că operația esențială în realizarea unei analize de conținut, de care depinde în cea mai mare măsură succesul cercetării, o constituie stabilirea categoriilor. Practic, lista categoriilor unei caracteristici este instrumentul tehnicii de față și faptul că unii autori îi atasează aceeași importanță ca și chestionarului în anchetă este pe deplin justificat. Cum operația de stabilire a categoriilor este uneori foarte dificilă, ea nu poate fi pe deplin realizată totdeauna înainte de începerea cercetării. Asta înseamnă că în anumite situații avantajul categoriilor se precizează, se constituie pe deplin doar pe măsură ce documentele sînt despuiate sau într-o fază pilot a cercetării, pe un subeșantion de texte.

Există, evident, foarte multe tipuri de categorii și nu se pot elibera rețete valabile pentru orice cercetare, tot așa cum nu se pot elabora chestionare universale. Voi menționa cîteva tipuri cu caracter pur ilustrativ:

(i) Dacă se studiază conținutul unor ziare, reviste, emisiuni de radio sau televiziune etc., categoriile vor surprinde subiectele tratate sau genurile de rubrici. De pildă, pentru programele de televiziune: filme, seriale, piese de teatru, emisiuni muzicale, varietăți, emisiuni de știri etc. Într-o eventuală analiză a revistei Viitorul social, articolele se pot clasifica după subramura sociologică în care se încadrează conținutul lor ș.a.m.d.

(ii) Un alt tip de categorii deseori folosit este cel de valori. Astfel, pentru a realiza, de exemplu, o analiză de conținut a manualelor de limba română din ciclul primar, ne putem propune drept scop să evidențiem valorile promovate în lecțiile respective. Analize similare sînt imaginabile asupra unui set de opere literare, emisiuni radio sau de televiziune, articole de ziar etc.

(iii) O serie specifică de categorii rezultă din abordarea însușirilor unor personaje: vîrsta, sexul, statutul social, trăsăturile fizice sau de caracter etc. În parte, aceste categorii coincid, după cum se vede, cu întrebările de clasificare din chestionar.

(iv) În fine, un alt exemplu tipic îl reprezintă categoriile de apreciere. Ele corespund variantelor sub formă de scală ordinală ale întrebărilor de opinie din chestionar. Caracteristici



cu asemenea categorii se pot construi doar atunci cînd în documente se găsesc aprecieri suficient de clare asupra unor fenomene sociale. De pildă, dacă ziarele iau atitudine față de un eveniment politic, se va încerca să se încadreze aceste poziții în categorii de genul: aprobare totală, aprobare parțială (moderată), indecizie, dezaprobare parțială (moderată), dezaprobare totală. Dat fiind faptul că textele nu sînt scrise ca să ne faciliteze nouă analiza de conținut, de cele mai multe ori este greu să realizăm un eventai prea larg de categorii de apreciere și ne vom mulțumi cu situația cea mai simplă: trihotomia favorabil, neutru (indecis), nefavorabil, termenii de apreciere putînd, evident, varia în funcție de problemă.

Pentru a-și îndeplini rolul important de le revine în sistematizarea materialului analizat, categoriile trebuie să fie:

- exhaustive (capabile să surprindă toate variantele posibile de apariție ale caracteristicii),

- exclusive (o unitate înregistrată într-o categorie să nu mai fie și în alta),

- obiective (să permită o clasificare a materialului care să depindă cît mai puțin posibil de subiectul care o realizează),

- pertinente (adică adecvate obiectivului urmărit de cercetare și conținutului documentelor).

Caracterul de exclusivitate nu este totuși general, așa cum nici variantele de răspuns la întrebările chestionarelor nu sînt totdeauna exclusive. Așa de exemplu, dacă efectuăm pomenita analiză a manualelor de citire, în cadrul unei lecții pot fi promovate mai multe valori, deci unitatea va trebui încadrată simultan în mai multe categorii.

Si cu aceasta am ajuns la a doua problemă fundamentală a analizei de conținut, și anume aceea de a preciza ce anume se urmărește în text, adică de a se specifica indicatorul empiric al categoriilor caracteristicii și de a se delimita spațiul unde acesta trebuie căutat. În cazul anchetei asemenea probleme nu se pun deoarece răspunsurile la întrebările chestionarului se adaptează, se dau în conformitate cu aceste întrebări. În analiza documentelor, dimpotrivă, întrebările se adaptează la informația conținută în documente și cum există o mare variabilitate a textelor trebuie să știm cum să depistăm existența unui răspuns la o problemă.



mă ce ne interesează. Apoi, în anchetă fiecare răspuns vine de la un individ uman, individ care se poate distinge net de ceilalți, ca entitate. În cazul documentelor lucrurile nu sînt nici în această privință atît de simple. În exemplele precedente, am sugerat cîteva unități cum ar fi numărul unei publicații, o emisiune de radio sau de televiziune, o lecție din manual, un articol dintr-o revistă științifică etc. Dar, pe de o parte, nu totdeauna documentele apar atît de clar fragmentate sau individualizate iar, pe de alta, fiecare unitate ca cele de mai sus poate fi descompusă în componente care, la rîndul lor, sînt apte de a fi luate în considerare ca unități distincte.

Va trebui deci stabilită, mai întîi, așa-numita unitate de înregistrare, adică entitatea purtătoare a informației căutate. Cea mai simplă unitate de înregistrare este, desigur, cuvîntul. Numai că limbajul natural este atît de complex încît - exceptînd situațiile în care studiul este de natură pur lingvistică, ceea ce e evident că pe noi nu ne interesează aici - rareori putem folosi acest indicator, din lipsa unei suprapunerii cît de cît acceptabile între cuvînt și noțiune. Vreau să spun că, de regulă, în sociologie nu ne interesează cuvîntul ca atare, ci semnificația acestuia și că în textele analizate putem întîlni cuvinte diferite cu semnificații identice sau același cuvînt, în alt context, cu semnificații diferite. Astfel, dacă în studiul unui discurs politic se urmărește apariția indicatorilor pentru categoriile de "patriotism" și "internaționalism", pentru prima dintre ele se pot lua ca unități de înregistrare termen ca: patrie, popor, națiune etc.

Se observă chiar și din acest singur exemplu cît de dificilă este alegerea cuvintelor capabile să reprezinte indicatori siguri pentru categoriile urmărite. Din acest motiv, studiile sociologie bazate pe cuvînt ca unitate de înregistrare nu sînt cele mai frecvente. Doar atunci cînd se pot identifica unele cuvinte cheie, cum este cazul în analiza textelor de propagandă politică, șansele de a duce la bun sfîrșit o lucrare sînt mai consistente.

Unitatea de înregistrare cea mai des folosită este tema, adică acel fragment din text care, prin ideile conținute, corespunde semnificației uneia sau alteia dintre categoriile stabilite. Spre pildă, dacă în una din lecțiile de citire, ca să relau un exemplu anterior, vom găsi descrierea unei bătălii pe care poporul român a



duș-o împotriva cotropitorului străin, atunci cu siguranță că unitatea respectivă trebuie înregistrată în categoria "patriotism", din evantaiul claselor de valori presupuse a fi promovate în textele respective. Se vede că nici depistarea ideilor, ca unități de înregistrare, nu este un lucru chiar atât de simplu, ba uneori operația se dovedește chiar mai dificilă decât în cazul cuvintelor sau noțiunilor. Reperarea unei teme într-un text este un act puternic influențat de trăsăturile celui care-l realizează și, din acest motiv, apreciez că tehnica analizei de conținut -chiar dacă sînt foarte riguros stabilite categoriile - nu poate fi realizată la cote calitative superioare decât de persoane cu multă experiență și cu un simț analitic foarte dezvoltat.

După ce sociologul stabilește cît mai concret posibil unitățile de înregistrare, adică indicatorii pe care trebuie să-i caute în text, urmează imediat precizarea ariei, a spațiului în care respectivii indicatori vor fi căutați. Cu alte cuvinte, universul documentelor trebuie împărțit în fragmente bine individualizate, astfel încît pe fiecare asemenea fragment să se urmărească dacă tema sau cuvîntul apare sau nu. Aceste unități poartă numele de unități de reperaj. Pe lîngă exemplele de unități de reperaj deja date (emisiunea de radio sau televiziune, număr sau articol de publicație, lecție de manual) se pot imagina încă multe altele. De pildă, dacă ne referim la un ziar și dorim să luăm o unitate de reperaj mai mică decît numărul sau articolul, putem merge cu fragmentarea textului în coloane, fracțiuni de coloane, alineate sau chiar fraze. La programele de radio sau televiziune, unitățile de reperaj mai mici decît emisiunea sînt, de regulă, de natură temporală (număr de minute de emisie). Astfel, în cadrul unor programe radiofonice de profil literar, cronometrînd intervale de 2,3 sau 5 minute, putem urmări pe acestea dacă numele unui scriitor este sau nu pronunțat, dacă sînt emise sau nu anumite idei etc.

În sfîrșit, dat fiind faptul că unitățile de reperaj nu sînt totdeauna egale ca întindere, în anumite situații va trebui să se țină seama și de dimensiunile acestora. Analizînd, de pildă numerele pe ultimii 20 de ani ai Revistei de filosofie și alegînd articolul ca unitate de reperaj, este clar că apariția unei teme într-o recenzie de o pagină nu poate fi identificată decât într-un mod foarte grosolan cu apariția aceleiași teme tratată însă într-un



studiu de, să zicem, 15 pagini. Prin urmare, aici se impune luarea în calcul și a dimensiunii unității de reperaj sau a întinderii pe care tema o prezintă. Pentru aceasta se introduce încă o noțiune: unitatea de numărare ca subdiviziune minimă pentru unitatea de înregistrare. În cazul anterior, unitatea de numărare ar putea fi pagina (pentru a nu coborî mai mult) și atunci fiecare temă ce apare într-un articol va fi luată în considerare de atâtea ori câte pagini are articolul respectiv.

Evident că în practică lucrurile sînt ceva mai complicate. Rareori o temă se întinde pe întreaga arie a unității de înregistrare și, din acest motiv, ea nu poate fi automat ponderată cu numărul unităților de numărare. Este posibil ca un studiu de 20 de pagini, luat ca unitate de înregistrare, să trateze o temă pe una sau două pagini, ceea ce înseamnă că tema respectivă va trebui numărată doar de una sau două ori și nu de 20. Sau, mai complicat, o temă poate apare pe mai multe pagini dar tratată de fiecare dată foarte sumar, într-o frază sau două (deci e departe de a ocupa întregul spațiu al paginilor respective). În astfel de situații, se va decide de la început asupra modului de lucru: sau se va lua ca pondere numărul concret de pagini pe care tema e identificată, indiferent de întinderea ei, sau se va evalua în pagini textul rezultat din cumulara fragmentelor specifice temei.

Pentru analiza de conținut a ziarelor, o unitate de numărare frecvent utilizată este cea cu care se poate măsura suprafața pe care se întinde textul, de exemplu centimetrul pătrat. Se ajunge astfel ca fiecare apariție de temă să fie evaluată în  $\text{cm}^2$ , ceea ce înseamnă practic o trecere de la numărare la măsurarea întinderii temei.

Modul concret de realizare a unei analize de conținut variază de la investigator la investigator și în funcție de natura documentelor cercetate. Atunci cînd avem de a face cu texte scrise - cazul cel mai frecvent - este recomandabil ca apariția indicatorilor categoriilor să fie notată pe marginea textului cu simbolurile alese pentru categorii. În felul acesta se poate verifica mai ușor că fiecare unitate de înregistrare a fost analizată, se facilitează construcția tabelelor de asociere între mai multe caracteristici și se poate mult mai simplu repeta analiza, deci verifica rezultatele obținute. Un exemplu cu două caracteristici e



ușor de imaginat în legătură cu analiza manualelor de citire despre care am mai pomenit. Fie prima caracteristică "valorile promovate", pe care o notăm cu A iar categoriile ei (valorile ce le vom distinge) cu  $A_1, A_2, A_3, \dots, A_s$  și a doua caracteristică "tipologia personajelor principale", notată cu B, avînd drept categorii (tipuri de personaje)  $B_1, B_2, B_3, \dots, B_t$ . Dacă luăm lecția drept unitate de înregistrare și numărare, în același timp, înseamnă că trebuie să urmărim pentru fiecare lecție dacă și care dintre categoriile de valori și personaje apar, notînd asemenea apariții cu simboluri  $A_i$  și /sau  $B_j$ . Astfel în dreptul unei lecții putem avea următoarea notație:  $A_2, A_5, A_7, B_3$ , corespunzînd faptului că am identificat aici trei categorii de valori iar personajul principal l-am clasat, după B, în categoria  $B_3$ . E posibil, evident, ca în dreptul unei lecții să nu putem trece nici un simbol pentru o caracteristică sau alta. De asemenea, se cere subliniat că atunci cînd un simbol de categorie este marcat în dreptul unei lecții el nu va mai fi marcat a doua oară chiar dacă ideea se repetă căci amales lecția, și nu un fragment al acesteia, ca unitate de înregistrare și numărare.

Rezultatul primar al analizei de conținut îl constituie construcția unor caracteristici statistice, de regulă calitative, adică a unor tabele simple de forma:

Categorii	$A_1$	$A_2$	$\dots$	$A_s$	Total
Nr. de apariții	$k_1$	$k_2$	$\dots$	$k_s$	$n$

ce se pretează la toate operațiile statistice ce vor fi descrise în partea a treia a cursului de față. O singură problemă specială se ridică aici, și anume aceea că numărul de apariții,  $n$ , nu coincide neapărat cu numărul unităților de înregistrare sau numărare, deoarece nu totdeauna categoriile sînt exclusive și, pe de altă parte, există unități în care nu se poate identifica niciuna din categoriile caracteristicii (situația analoagă, formal, nonrăspunsului la o întrebare din chestionar). De aceea, cifrele relative, în speță procentele, se calculează la mai multe totaluri: al numărului de apariții, al numărului de unități de înregistrare etc.



De pildă, dacă setul de manuale analizate cuprinde 100 de lecții, numărul de apariții al valorii "patriotism" este 50 (această valoare e promovată în 50 de lecții) iar numărul total de apariții pentru întreg setul de valori este, să zicem 250, atunci frecvența 50 se poate raporta fie la totalul de 100 (50% din lecții promovează valoarea respectivă), fie la 250 (20% dintre apariții revin valorii "patriotism").

Am ales mai sus spre exemplificare niște cazuri și cifre fictive deoarece exemple de analiză de conținut în literatura noastră sociologică sînt puține și nu foarte semnificative pentru scopurile didactice aici urmărite. Aș aminti totuși cartea Sociologie românească contemporană (1983), în care sînt expuse, printre altele, și rezultatele unei ample cercetări a scrierilor sociologice din perioada postbelică, cercetare ce poate fi asimilată analizei de conținut. Unitățile de înregistrare au fost lucrările sociologice, dar "calibrul" lor foarte diferit i-a condus pe cercetători să grupeze materialul pe 6 dimensiuni: volume, culegeri de studii, articole, cercetări, teze de doctorat și manuale. Caracteristica utilizată a fost tematica lucrărilor, clasificată în 21 de categorii. Tabelul frecvențelor, după care s-a întocmit și ierarhia categoriilor, apare la pagina 103 a lucrării citate.



Cap. VI.

EXPERIMENTUL

În cadrul activității de cercetare, atunci când este vorba de științele empirice și, mai ales, de cele ale naturii, experimentul este considerat, de regulă, ca fiind metoda cea mai consistentă de cunoaștere. Această poziție privilegiată a experimentului în rândul metodelor prin care omul reușește să explice fenomenele naturii, prin care se descoperă, se evidențiază și se confirmă legăturile cauzale dintre fenomene, nu putea să nu atragă atenția și cercetătorilor din domeniul științelor sociale.

Dar specificul acestor discipline - și al sociologiei, mai cu seamă - ridică de la bun început o serie de obstacole în calea utilizării metodei experimentale, provocând astfel numeroase discuții asupra posibilității însăși a experimentării în științele umane, discuții care umplu mii de pagini în lucrările de natură metodologică și epistemologică. Dacă cineva și-ar da silința, probabil că ar descoperi inimaginabil de multe poziții, caracterizate prin nuanțe și subtilități deosebite, care merg pe o scală ce începe cu negarea oricărei posibilități de experimentare în sociologie și se sfârșește cu concepția care pune un semn de egalitate (în ceea ce privește utilitatea, posibilitățile de cunoaștere și de realizare) între experimentul sociologic și cel din științele fizice.

Chiar dacă lăsăm la o parte de la bun început filozofii și sociologii care consideră realitatea socială complet distinctă (și chiar opusă) față de cea fizică, cel puțin din punctul de vedere al posibilităților de cunoaștere, deci pe cei care introduc o ruptură între metodele de cunoaștere din științele naturii și cele socio-umane, lăsînt, prin urmare, la o parte pe acei care, în mod normal, prin poziția lor filozofică, nu pot accepta experimentul ca un mijloc de cunoaștere a vieții sociale și referindu-se numai la cei care acceptă, implicit sau explicit, o unitate a metodelor de cunoaștere științifică, vom găsi, referitor la experiment pozițiile extreme amintite mai sus. Restrîngînd și mai mult aria, și făcînd apel numai la gînditori marxisti, vom identifica practic aceeași diversitate de poziții. Astfel T.Nicola (1973, pp.58-59)



examinând oțeva opinii exprimate într-o dezbaterē din literatura sovietică, identifică trei puncte de vedere:

1. Experimentul este posibil numai în cunoașterea fenomenelor naturii și alte tehnicii.

2. Experimentul poate fi utilizat în studierea fenomenelor sociale, dar metoda este esențialmente diferită de cea din științele naturii.

3. Experimentul se poate folosi cu succes în cercetarea fenomenelor sociale, adevind metoda la specificul acestora.

Deși mărturisesc că nu-mi este prea clară distincția dintre pozițiile 2 și 3, bănuiesc totuși că este vorba de o gradație a atitudinilor față de experimentul social.

Pentru a nu mai zăbovi prea mult în discutarea acestor poziții, voi mai ilustra doar una ce mi se pare deosebit de semnificativă. Este vorba de concepția lui E. Durkheim, sociologul care a militat cu atîta forță pentru a impune necesitatea privirii faptului social ca un lucru, deci pentru a apropia cunoașterea sociologică de cea din științele naturii. Imi permit să dau un citat mai amplu, de fapt începutul capitolului VI din Regulile metodei sociologice (1974) spre a înțelege exact poziția lui în problema ce o discutăm aici (sublinierile îmi parțin):

"N-avem decît un mijloc pentru a demonstra că un fenomen este cauza altuia, acela de a compara cazurile în care ele sînt simultan prezente sau absente și de a căuta dacă variațiile pe care le prezintă, în diverse combinații de circumstanțe, dovedesc că unul depinde de celălalt. Cînd ele pot fi produse în mod artificial după placul observatorului, atunci metoda este experimentală propriu-zisă. Cînd, dimpotrivă, producerea faptelor nu este la dispoziția noastră și nu putem decît să le aproximăm așa cum s-au produs spontan, atunci metoda care se întrebuintează este aceea a experimentării indirecte sau metoda comparativă.

Am văzut că explicația sociologică constă exclusiv în a stabili raporturi de cauzalitate, fie că e vorba de a ratașa un fenomen la cauza lui sau, dimpotrivă, o cauză la efectele sale utile. Pentru că, pe de altă parte, fenomenele sociale scapă evident acțiunii operatorului, metoda comparativă este singura care se potrivește sociologiei" (p.164)



Așadar, este limpede, cred, deși Durkheim mai folosește uneori în continuare termenul "de experiment" ca sinonim pentru metoda comparativă, că pentru sociologul francez metoda experimentală "propriu-zisă" este apanajul exclusiv al științelor naturii.

Problema care se ridică și consider că trebuie rezolvată în acest moment este cea de a explica posibilitatea unei diversități atât de mari de opinii, care, cum spuneam, ating poziții extreme. După părerea mea lucrurile se explică relativ simplu, în funcție de două elemente esențiale:

1. ce definiție adoptăm pentru sociologie;
2. ce definiție adoptăm pentru experimentul în sociologie.

Referitor la prima problemă, se poate ușor constata că dacă se ia pentru sociologie un obiect mai larg, în sensul în care unele definiții vorbesc despre "societatea în ansamblul său", "sistemul social global" etc., atunci este firesc că posibilitatea experimentării este substanțial redusă. Dacă, dimpotrivă, se acceptă ca aparținând sociologiei fenomene sau procese ce caracterizează grupuri mai restrânse sau acțiuni sociale foarte particulare, atunci este iarăși evident că o serie de obstacole "materiale" în calea experimentării cad.

În al doilea rând, așa cum vom vedea în continuare, dacă se identifică experimentul social cu modelul ideal de experiment din științele naturii, atunci șansele de a experimenta în sociologie - și chiar în alte științe sociale - sînt mici. Dimpotrivă pe măsură ce se renunță la condițiile de definiție a experimentului, deci dacă se acceptă definiții mai puțin "tari", atunci posibilitățile de a realiza asemenea experimente cresc.

Avînd în vedere acest din urmă aspect și pentru a putea oferi un ghid folositor pentru a înțelege multitudinea de atitudini referitoare la experiment, am considerat necesar să expun pe scurt esența metodei experimentale.

### 1. Ce este experimentul ?

Posibilitățile de definire a experimentului sînt extrem de numeroase, fapt care influențează, cum spuneam și atitudinile față de această metodă. Psihosociologul S.Chelcea (1982, p.95) amintește că E.Greenwood, cărui îi datorăm o lucrare fundamentală, Experi-



mental Sociology. A study in Method, apărută în 1945, identifică în literatura sociologică și psihosociologică din primele patru decenii ale secolului nostru peste 100 de definiții ale experimentului.

În ceea ce mă privește, voi încerca să relievez esența metodei experimentale, așa cum ea s-a constituit în științele naturii, reținând acele caracteristici fundamentale pe care trebuie să le îndeplinească orice experiment indiferent de domeniul său de realizare.

Prima observație care se cuvine făcută este aceea că experimentul este superior observației ca eficiență cognitivă pentru că el implică toate rigorile acesteia, adăugând și altele suplimentare. Mă gândesc în special la faptul că, la fel cu observația, experimentul presupune urmărirea conștientă și sistematică a manifestării fenomenelor studiate, precum și înregistrarea manifestărilor respective. Această urmărire poate fi executată cu (iar înregistrearea facilitată de) diferite instrumente și aparate, precum microscopul, ampermetrul etc. Mi se pare foarte limpede că un cercetător din câmpul științelor naturii include asemenea activități în orice situație experimentală.

Ce aduce în plus experimentul? După părerea mea, două lucruri, ambele la fel de importante și deci indispensabile:

(i) Întâi, experimentul presupune, spre deosebire de observație, intervenția activă a subiectului cunoscător în desfășurarea fenomenului studiat. Această intervenție activă se manifestă fie prin introducerea, provocarea unui factor nou (unor factori noi), inexistent (inexistenți) în realitatea studiată, fie prin dirijarea variației unui factor existent (unor factori existenți).

(ii) În al doilea rând, experimentul presupune controlul unui număr de factori. Termenul de control trebuie înțeles aici într-un sens ceva mai restrâns decât îl are în limba engleză. Într-adevăr, se poate înțelege prin control însăși acea intervenție de introducere sau variație a unui factor. Pentru a deosebi aspectele prezentate la (i) și (ii), prin control voi înțelege numai o acțiune limitativă, de îngrădire a manifestării altor factori, cu scopul de a fi siguri că singura schimbare esențială petrecută în complexul studiat este cea provocată de cercetător.



Prin urmare, experimentul presupune o intervenție a subiectului cercetător în dublu sens: de producere, de provocare a unor manifestări și de constrângere, de limitare a altora. De exemplu, prin control în fizică se poate înțelege fie izolarea fenomenelor de unii factori perturbatori (cazul căderii corpurilor în camere cu vid), fie menținerea constantă a unor factori, cum ar fi presiunea, temperatura etc. Cercetătorul din științele naturii are, de regulă, la dispoziție instrumente și aparate adecvate pentru realizarea ambelor tipuri de acțiuni.

Pentru a circumscrie mai exact experimentul ca metodă științifică de cunoaștere, mai menționez că și aici, ca și în cazul observației, acțiunea respectivă trebuie să aibă un caracter eminamente de cunoaștere, în două sensuri: atât prin premisele sale, cât și prin finalitatea ei. Ambele sublinieri sînt necesare pentru a evidenția, pe de o parte, faptul că orice experiment științific pornește de la o ipoteză (sau mai multe), care derivă dintr-o teorie, mai mult sau mai puțin elaborată și conștientizată, a fenomenelor cercetate. Pe de altă parte, finalitatea științifică este o caracteristică ce evidențiază diferența dintre experiment și tot soiul de "experiențe" care sînt atât de des întîlnite în viața socială.

După părerea mea, o metodă științifică ce se dorește experiment trebuie să tindă, indiferent de domeniul de aplicare, spre realizarea tuturor condițiilor amintite mai sus.

De multe ori prin experiment se înțelege numai intervenția activă, menționată la punctul (i). Nu e greu de văzut că o asemenea concepție duce în fapt la identificarea experimentului sociologic cu aproape orice acțiune socială, cunorice activitate practică transformatoare. Confuzia cea mai gravă ce se poate face (și se face foarte des) în sensul acesta este între experiment și activitatea de conducere socială. Orice act de conducere înseamnă o modificare a socialului; orice dispoziție, măsură, reglementare juridică etc. este o intervenție în desfășurarea fenomenelor sociale, cu consecințe mai evidente sau mai ascunse, mai apropiate sau mai îndepărtate, mai previzibile sau mai neașteptate etc., dar pentru că tuturor acestora le lipsește, în genere, controlul factorilor și finalitatea de cunoaștere, ele nu pot fi socotite experimente. De aici nu trebuie să se tragă concluzia că experimentul



n-ar putea sau n-ar trebui să aibă aplicații practice. Aceasta este însă o însușire care nu este esențială, definitorie pentru actul științific ca atare. Desigur că știința se dezvoltă pe seama activității practice, fiind influențată de solicitările acesteia, că știința are finalități foarte practice ș.a.m.d., dar aplicațiile practice ale științei nu trebuie confundate cu știința însăși.

Alteori se renunță în definirea experimentului tocmai la intervenția activă (i). Motivul nu este greu de înțeles și se bazează pe ideea că cercetătorul poate profita de anumite schimbări care se produc fără intervenția sa. Un cutremur, o inundație sau alte calamități naturale pot produce schimbări în viața socială, care se cer studiate. Cu toate acestea, nu cred că asemenea situații pot fi numite experimente. Mai întâi, pentru că nici în științele naturii nu se folosește acest termen când se studiază modificările produse de natura însăși. Așa de pildă, nu am sesizat niciodată că astronomii ar numi eclipsele de soare sau de lună experimente, cu toată înfrigurarea și interesul cu care sînt așteptate evenimentele respective. În al doilea rînd, față de științele naturii, unde schimbările sînt în general mai lente, stabilitatea sistemelor mai mare, științele sociale au, de regulă, de a face cu transformări mult mai evidente și atunci, practic, orice studiu care urmărește un sistem în diacronia sa ar putea fi numit experiment.

Probabil însă că discuția aceasta ar putea fi cu mult mai folos continuată, pentru că argumentele mele vor fi mai ușor de înțeles, după ce analizăm, chiar dacă foarte schematic, structura situațiilor experimentale.

## 2. Schema logică a experimentului

La ora actuală, schemele experimentale au ajuns la complexități și subtilități nebănuite de omul de rînd, tocmai cu scopul de a se realiza cît mai fidel cerințele din definiție, mai ales cea referitoare la controlul factorilor. Dar oricît de sofisticate ar fi formele de experimentare, ele pornesc de la o schemă relativ simplă, pe care o voi expune în cele ce urmează.

Ne vom plasa în cazul elementar al testării legăturii



dintre două fenomene, din care unul, notat cu  $y$ , este considerat a fi efectul sau consecința celuiilalt, notat cu  $x$ . Pentru simplificarea limbajului, vom numi pe  $x$  variabilă independentă iar pe  $y$  dependentă.

În afara acestor factori, mai trebuie avute în vedere încă două mulțimi de variabile:

- grupul, pe care-l notăm cu  $A$ , ce cuprinde toți acei factori care, pe baza teoriei sau a cunoștințelor noastre empirice, considerăm că pot avea o influență semnificativă asupra lui  $y$  sau asupra relației dintre  $x$  și  $y$ ;

- grupul, pe care-l notăm cu  $B$ , ce cuprinde restul factorilor, a căror influență o putem considera nulă sau, în orice caz, de un ordin de mărime net inferior celui al factorilor din  $A$ .

În continuare, experimentul mai presupune observarea situațiilor în minimum două momente de timp:

-  $t_0$ , momentul inițial, anterior realizării acțiunii modificatoare;

-  $t_n$ , momentul final, după realizarea acțiunii, când se evaluează efectul lui  $x$  asupra lui  $y$ .

Se ajunge la următoarele două scheme:

	$t_0$	$t_n$
(I)	$A_0, B_0, y_0$	$x, A_n, B_n, y_n$
(II)	$x_0, A_0, B_0, y_0$	$x_n, A_n, B_n, y_n$

Primul caz corespunde situației când variabila  $x$  este efectiv introdusă, factorul respectiv neexistînd la momentul  $t_0$  al doilea, când variabila  $x$  suferă o transformare, trecînd de la  $x_0$  la  $x_n$ . Pentru ca situațiile noastre să îndeplinească și a doua condiție fundamentală a experimentului, este nevoie ca factorii din complexul  $A$  să fie controlați, adică acțiunea lor să nu se modifice între  $t_0$  și  $t_n$ . Vom nota prescurtat aceasta prin:

$$A_0 = A_n$$

subliniind că semnul egal indică aici o identitate de stare și nu o egalitate în sens matematic a două mărimi.



Din situația (I) va rezulta deci că, dacă toți factorii  $\underline{A}$  sînt constanți, schimbarea lui  $\underline{y}$  de la starea  $\underline{y}_0$  la  $\underline{y}_n$  se datorează introducerii factorului  $\underline{x}$ . Vom nota această schimbare cu

$$\Delta_y = y_n - y_0$$

unde semnul minus poate însemna efectiv o diferență matematică, dacă factorul  $\underline{y}$  este măsurabil, sau numai o diferență calitativă de stări, dacă nu este măsurabil.

În situația (II) se observă că variația lui  $\underline{x}$ , de la starea  $\underline{x}_0$  la  $\underline{x}_n$ , ce poate fi simbolizată prin:

$$\Delta_x = x_n - x_0$$

produce variația lui  $\underline{y}$ :

$$\Delta_y = y_n - y_0$$

Desigur că în acest caz putem urmări relația dintre  $\underline{x}$  și  $\underline{y}$  nu numai în cele două momente, inițial și final, ci și pe tot intervalul de timp cît durează experimentul. Astfel, va rezulta o relație de dependență între  $\underline{x}$  și  $\underline{y}$ , ce poate fi pusă sub formă unui tabel al observațiilor:

Stările lui  $\underline{x}$ :  $x_0 \quad x_1 \quad x_2 \quad \dots \quad x_n$

Stările lui  $\underline{y}$ :  $y_0 \quad y_1 \quad y_2 \quad \dots \quad y_n$

dependență care, în cazul cînd  $\underline{x}$  și  $\underline{y}$  sînt măsurabile, poate lua forma unei funcții matematice:

$$y = f(x)$$

Se mai observă că cele două cazuri, (I) și (II), corespund la două dintre cunoscutele "canoane" ale lui J.St.Mill. Primul descrie metoda diferenței, dacă prin fenomenul efect înțelegem tocmai variația lui  $\underline{y}$ , iar al doilea, metoda variațiilor concomitente. Într-adevăr, formulele (I) și (II) pot fi scrise:



$$(I) \quad \begin{matrix} t_0 & & t_n \\ A_0, B_0, y_0 & & x, A_0, B_n, y_0 + \Delta y \end{matrix}$$

$$(II) \quad \begin{matrix} x_0, A_0, B_0, y_0 & & x_0 + \Delta x, A_0, B_n, y_0 + \Delta y \end{matrix}$$

Din (I) rezultă că  $x$  produce pe  $\Delta y$  iar din (II) că  $\Delta x$  generează pe  $\Delta y$ . Tot din schematizarea de mai sus apare ca o concluzie firească faptul că metoda diferențelor este un caz particular al celei a variațiilor concomitente, și anume atunci cînd  $x_0 = 0$  și cînd pentru (II) se au în vedere numai momentele  $t_0$  și  $t_n$ .

Așadar, din cele patru canoane propuse de Mill, două se dovedesc adecvate situațiilor experimentale, rezultînd deci că celelalte, metoda concordanței și cea a reziduurilor, sînt specifice observației. De aici nu rezultă însă că cele care se potrivesc experimentului nu pot fi folosite în cazul observației, ci dimpotrivă că ele au virtuți cognitive superioare în observație tocmai pentru că reflectă logica experimentului. Pentru ca metoda diferenței sau cea a variațiilor concomitente să se aplice în cazul observației este nevoie de întrunirea a două condiții:

-factorii  $A$  să rămînă neschimbați;

--observatorul săi aibă, în plus, șansa de a asista la aparitia neprovocată a fenomenului  $x$  (pentru cazul I) sau la variația acestuia (pentru II)-

Prin urmare, aplicarea în sine a canonului diferenței sau al variațiilor concomitente nu înseamnă experiment; se ajunge la experiment numai în măsura în care introducerea sau variația factorilor independenți este provocată, asigurîndu-se în același timp controlul factorilor  $A$ . Renunțarea la condiția care cere manipularea variabilei independente are drept consecință, prin slăbirea definiției, lărgirea sferei noțiunii de experiment pînă la cuprinderea în ea a unor proceduri care nici pe departe nu mai seamănă cu experimentul tradițional. Iar denumirea unor astfel de proceduri cu termenii de "experimente indirecte" sau "semiexperimente" nu face decît să creeze confuzie, mai cu seamă atunci cînd unii autori le acordă, implicit sau explicit, statutul de experiment.



Un exemplu de asemenea experimentare indirectă este așa-numita "experimentare ex post facto". "Ea constă, spune M. Duverger (1964), în cercetarea înălțurii factorilor care conduc la o situație dată, după ce aceasta s-a produs" (p.367). Voi schița acest așa-zis experiment tocmai pentru a se vedea cât de departe ne aflăm de o situație experimentală propriu-zisă. De fapt, există aici două tipuri de abordate : de la cauză spre determinarea efectelor sale și de la efect spre determinarea cauzei. Pentru ilustra-re, voi cita chiar exemplele date de Duverger, provenind din cerce-tările sociologilor americani.

Primul exemplu urmărește stabilirea efectului pe care in-strucția îl are asupra reușitei sociale. Un cercetător a pornit de la un lot de 1.194 subiecți, care, cu 9 ani mai înainte, au absolvit o aceeași școală . Dintre ei o parte și-au continuat stu-diile câțiva ani iar alții au intrat în câmpul muncii (variabila independentă). Fiecare subgrup a fost astfel redus încât loturile rămase să fie perfect echilibrate din punctul de vedere al originii sociale, al nivelului intelectual, al naționalității de origine etc. (controlul factorilor). Până la urmă s-au ajuns la câte 145 de indivizi în fiecare grup și s-a evaluat reușita socială a indivi-zilor din cele două loturi (variabila dependentă).

Al doilea exemplu - pentru ilistrarea demersului invers - se referă la o cercetare făcută în Minnesota, avînd ca obiectiv compararea a 50 delinvenți cu 50 de nondelinvenți, identici cu pri-mii din punctul de vedere al unității școlare absolvite, al ni-velului de inteligență, al situației părinților etc. S-a încercat urcarea în trecut spre a descoperi cauza delinvenței.

Asemenea situații "experimentale" se pot imagina foarte ușor, dar mie mi se pare că ele nu se deosebesc de simpla observa-ție (în sensul în care Durkheim înțelegea metoda comparativă), decît poate printr-o grijă deosebită purtată controlului factori-lor. Dar oare se mai practică astăzi observații în care contro-lul factorilor să nu stea în atenția cercetătorului ? Iar fiindcă sociologia este o disciplină care apelează cel mai des la o abor-dare statistică a fenomenelor și proceselor studiate, nu este lîp-sit de interes de subliniat că o direcție fundamentală în dezvoltarea



tarea statisticii în ultimul secol, a fost tocmai aceea care a furnizat instrumente eficiente de control statistic al factorilor. E suficient să amintesc aici de analiza corelațiilor parțiale și multiple, analiza de regresie, analiza path și de dependență, analiza multivariată a caracteristicilor calitative etc., care toate au ca scop, în ultimă instanță, să "purifice" relația dintre două variabile de acțiunea perturbatoare a altor factori. De fapt orice tabel statistic, care asociază sau corelează simultan cel puțin trei variabile, poate fi interpretat în același scop.

În concluzie, voi afirma că o cercetare nu are caracter de experiment decât dacă îndeplinește condițiile definitorii precizate mai sus. Dacă nu intervine activ în generarea factorului independent, atunci cercetătorul procedează doar la o observație, care, chiar dacă e mai riguroasă prin control, tot observație rămîne. Dacă nu realizează un control suficient și nu urmărește scopuri științifice (aceste două elemente fiind, de regulă, legate), atunci activitatea sa se confundă cu un tip sau altul de practică socială.

### 3. Particularitățile experimentării în sociologie și în științele socio-umane, în general

Unii autori încearcă, așa cum s-a văzut, să găsească deosebiri esențiale între experimentul din științele sociale și cel din științele naturii; alții deosebiri mai puțin importante. Din păcate, când se vorbește despre astfel de diferențe rareori se precizează pe ce plan se consideră cele două experimente ca fiind diferite: pe cel logico-epistemologic sau pe cel al realizării practice. Altfel spus, trebuie precizat, când vorbim de deosebiri, dacă experimentul social prezintă particularități specifice în structura sa esențială, deci ar necesita o definiție deosebită, sau experimentul fiind unul în orice știință, atunci când ne referim la realitatea socio-umană apar dificultăți în aplicarea experimentului, de unde pot rezulta eventual anumite forme specifice de realizare, care nu afectează însă cu nimic structura sa generală.

Poziția mea cred că a reieșit deja suficient de clar din



cele spuse pînă aici: consider că, din punctul de vedere al structurii sale logice și al trăsăturilor definitorii, experimentul este același indiferent de domeniul de aplicare. Diferențieri se pot ivi doar în modalitățile practice de realizare a acestei acțiuni de cunoaștere. Dacă prin definiție experimentul este ceva în fizică și altceva în sociologie, atunci s-ar putea ridica - și se ridică - pe bună dreptate întrebarea de ce asemenea metode trebuie să se mai numească ambele experimente.

Pe de altă parte, se cuvine menționat că îndeplinirea riguroasă a tuturor condițiilor (mai ales cea privind controlul factorilor) este mai degrabă un ideal decît o realitate, chiar în științele naturii cu obiectul cel mai simplu, cum este fizica. Adeseori, ca în multe alte probleme de metodologie, se face greșeala de a se identifica experimentul în științele naturii cu idealul experimentării, contrapunînd apoi acestuia experimentul social. Adevărul este că și științele naturii se deosebesc în funcție de măsura în care reușesc să se apropie de acest ideal. Nu toate științele beneficiază în aceeași măsură de posibilitatea provocării unor fenomene, de posibilitatea controlului diferiților factori sau a integrării rezultatelor într-o textură teoretică solidă. Rezultă, prin urmare, că deosebirea dintre observație și experiment nu este una absolută. Există, cum spuneam, observații care tind să înglobeze cît mai mult cerințele experimentului, așa cum există multe experimente imperfecte, din punctul de vedere al realizării lor efective. Aceasta nu înseamnă însă că trebuie să ștergem orice graniță între observație și experiment. Aceasta nu înseamnă că avem dreptul să decretăm experiment orice observație mai sofisticată numai pentru a putea, chipurile, depăși condițiile unei științe cu slabe posibilități experimentale. De altfel, există științe foarte venerabile, ca astronomia, care nu apelează deloc la metoda experimentală și prin aceasta nimeni nu le pune la îndoială nivelul de științificitate al propozițiilor.

Cît privește sociologia - și în general științele sociale - există dificultăți metodologice generale în abordarea obiectului lor, despre care am vorbit la momentul potrivit. Nu le voi mai repeta aici, ci voi încerca să menționez forma concretă în care ele se prezintă atunci cînd urmărim realizarea unui experiment. Pentru o oarecare sistematizare, o să mă ghidez după elementele



utilizate în definirea acestei metode.

Mai întâi, se impun atenției o serie de obstacole legate de intervenția activă în viața socială, de introducerea unor evenimente noi în viața unor comunități umane. Pentru omul de știință dificultățile respective cresc, poate mai mult decât proporțional, pe măsura ridicării de la micro la macrosocial. Asemenea intervenții întâmpină obstacole, să le spunem, materiale, în sensul că cercetătorul nu are pur și simplu puterea de a interveni; este cât se poate de evident că, în măsura în care se implică în experiment grupuri cât mai mari de oameni, el trebuie să facă apel la organele conducătoare, politice și administrative, și să primească nu numai aprobarea dar și sprijinul lor efectiv pentru realizarea mutațiilor dorite. Un alt gen de obstacole îl reprezintă cele morale: una este să experimentezi pe cobai și alta pe pielea semenilor tăi. Chiar dacă se poate presupune cu mult temei că experimentul nu va influența semnificativ viața celor investigați, elementarele reguli ale eticii ne cer să aducem la cunoștința și să primim consimțământul acestora pentru realizarea experimentului.

Dar cele mai serioase greutăți în calea experimentării în științele sociale apar la nivelul controlului factorilor. Se pot distinge și aici mai multe planuri pe care se manifestă aceste dificultăți.

Se cuvine subliniată, în primul rând, complexitatea fenomenelor sociale, interdependențele extrem de numeroase de aceeași intensitate sau de intensități apropiate, ceea ce nu se întâmplă în aceeași măsură în științele fizice. Aș reaminti celebrul experiment asupra căderii corpurilor în camera cu vid. Atunci când "controlăm" densitatea atmosferică, în sensul că eliminăm aerul din câmpul experienței, vom controla singurul factor cu adevărat perturbator care influențează căderea corpurilor; în lipsa acestui factor, legea rămîne și se manifestă în toată puritatea ei. Poate că mai sînt și alți factori care, la o analiză extrem de minuțioasă, ar putea manifesta o influență asupra căderii corpurilor, dar influența acestora se va dovedi de un ordin de mărime atît de mic încît se poate ca ușurință neglija. Să încercăm să găsim asemenea situații în societate! Cîți factori trebuie controlați pentru a ajunge la explicarea delinvenței, a mobilității sociale, a



migrației etc. ? Si este evident că, pe măsură ce numărul factorilor ce trebuie controlați este mai mare, cresc și dificultățile respectivului control.

Pe de altă parte, în științele naturii se utilizează aparate, instrumente care permit izolarea fenomenului asupra căruia experimentăm. În felul acesta controlul se poate realiza eficient și, la rândul său, poate fi controlat, fiind astfel siguri că realmente are loc controlul dorit. Izolarea în științele sociale nu beneficiază, de regulă, de asemenea instrumente valide și ca atare se poate realiza doar în mod cu totul precar, rămânând de multe ori doar la nivelul intențiilor, fără a putea fi verificată.

În al treilea rând, este vorba de existența unui factor ce nu se întâlnește în științele naturii, factor care ridică dificultăți absolut inedite în fața cercetătorului social, mai cu seamă în legătură cu controlabilitatea lui: mă refer la realitatea că oamenii știu că sînt supuși experimentului, deci că ei sînt subiecți ai investigației. Voi nota acest factor cu C. Acțiunea lui C, în anumite situații experimentale, poate fi atît de importantă încît să estompeze aproape total raportul dintre factorii studiați în cadrul experimentului.

Un exemplu mai mult decît semnificativ în acest sens îl constituie unul dintre celebrele experimente de la Hawthorne, desfășurate de Elton Mayo și echipa sa la uzina Western Electric la sfîrșitul anilor '20. E vorba de experimentul privind influența iluminării locului de muncă asupra randamentului (De fapt acest experiment a început înainte de sosirea lui E. Mayo la Hawthorne, dar el este inclus, de regulă, în șirul de experimente desfășurate în cadrul programului respectiv). S-au luat două grupuri de muncitori, unul la care iluminatul a rămas constant și altul la care a variat. La ambele echipe productivitatea muncii a crescut simțitor; la cea cu iluminare variabilă, productivitatea a sporit atît la mărirea gradului de iluminare cît și la diminuarea acestuia pînă la limita vizibilității. Este deci evident că factorul care a dus la creșterea productivității a fost însăși conștiința indivizilor că sînt observați, că se găsesc într-o situație cu totul specială. Deci relația de dependență presupusă între  $y$  (productivitate) și  $x$  (iluminat) a rămas finalmente ascunsă datorită factorului C.



S-ar putea desigur înmulți la nesfârșit exemplele care să confirme aprecierea că dificultățile controlului sînt foarte accentuate în experimentele din științele socio-umane, dar consider că lucrurile sînt deja suficient de clare. Se ridică atunci firesc întrebarea dacă această cerință fundamentală a experimentului - controlul - este realizabilă (sau realizabilă într-o măsură satisfăcătoare) pentru a putea vorbi de experimente și în acest cîmp al cunoașterii. Răspunsul este afirmativ fiindcă, din fericire, inventivitatea omului este fără limite și, în cazul de față, s-a găsit o procedură satisfăcătoare de control fără a se acționa direct asupra factorilor controlați. Este vorba despre experimentul cu grup (grupuri) de control, în care controlul se realizează în mod indirect. Ar fi de precizat că această tehnică experimentală nu este specifică științelor sociale; ea se folosește în multe sectoare ale științei, peste tot unde controlabilitatea este redusă. Biologia, științele agricole și cele medicale sînt numai cîteva exemple de asemenea discipline pentru care experimentul cu grup de control este o metodă curent întrebuințată.

Schema logică a experimentului de tip (I), cu grup de control este:

$$\begin{array}{l} \text{G:} \quad A_0, B_0, y_0 \quad \quad \quad x_n, A_n, B_n, y_n \\ \text{G':} \quad A_0, B'_0, y_0 \quad \quad \quad A_n, B'_n, y'_n \end{array}$$

unde prin G am notat grupul experimental iar prin G' grupul de control.

Grupul de control G', se alege de așa manieră încît să coincidă în cea mai mare măsură cu grupul experimental, ceea ce înseamnă aceeași stare a variabilelor cuprinse în A și aceeași valoare a lui y la momentul inițial. De asemenea, se mai face supoziția că, între t<sub>0</sub> și t<sub>n</sub> setul factorilor A urmează aceeași variație în ambele grupuri, trecînd de la starea A<sub>0</sub> la A<sub>n</sub>.

Se vede deci că, astfel ales grupul de control, într-o asemenea situația experimentală, nu mai este nevoie să se pună condiția:



$$A_n = A_0$$

care, așa cum am văzut, creează cele mai mari obstacole în calea experimentului social. Într-adevăr, chiar dacă factorii  $A$  variază între cele două momente de timp, din ipotezele menționate mai sus rezultă că efectul introducerii factorului  $x$  numai în grupul experimental este marcat în diferența de stări la care ajunge factorul dependent,  $y$  în cele două grupuri la momentul final. Deci, formal, efectul lui  $x$  va fi:

$$y_n - y'_n$$

și nu

$$y_n - y_0$$

ca în cazul experimentului fără grup de control. În adevăr, punînd

$$y_n - y_0 = y_n - y'_n + y'_n - y_0$$

se vede că schimbarea variabilei în grup experimental ( $y_n - y_0$ ) se descompune în suma dintre variația lui  $y$  în grupul de control ( $y'_n - y_0$ ) și diferența valorilor rezultate în cele două grupuri ( $y_n - y'_n$ ). Admițînd că variația în grupul de control se datorează trecerii factorilor  $A$  din starea  $A_0$  în  $A_n$  și că aceeași modificare se produce și în grupul experimental, rezultă că efectul introducerii noului factor  $x$  este  $y_n - y'_n$ .

Un exemplu simplu de astfel de experiment poate fi luat din domeniul pedagogiei. Să presupunem că un profesor dorește să evalueze eficiența unei metode noi de instrucție. Va alege două clase de elevi cît mai apropiate din punctul de vedere al structurii și cu rezultate identice la disciplina de studiu respectivă. În o clasă va utiliza nouă tehnologie pedagogică iar la cealaltă va rămîne la vechea metodă folosită. Dacă cele două colectivități au fost corect alese și pe parcursul experimentului (de pildă un an școlar) nici un grup nu a fost supus influenței vreunui alt factor specific, atunci diferența dintre rezultatele școlare ale celor două clase poate fi pusă pe seama factorului nou introdus.

Pentru experimentul de tip (II), schema formală, în cazul cînd se folosește un grup de control, va fi:



$t_0$

$t_n$

G:  $x_0, A_0, B_0, y_0$   $x_n, A_n, B_n, y_n$

G':  $x_0, A_0, B'_0, y_0$   $x_n, A_n, B'_n, y'_n$

Diferența dintre cele două grupuri este aici marcată de faptul că în cel de control factorul  $x$  rămâne constant, pe cînd în cel experimental el este variabil. Efectul variației lui  $x$  va fi și în acest caz diferența dintre stările finale ale lui  $y$  în cele două grupuri.

Se înțelege că formele prezentate mai sus reprezintă schemele cele mai simple de experiment, care surprind esența procedurii științifice analizate aici. Ele se pot "complica" în funcție de diferite elemente concrete care intervin în aplicarea lor în practică. De exemplu, în cazul (I), dacă nu se poate asigura identitatea stărilor lui  $y$  la momentul inițial în cele două grupuri, deci dacă vom avea în G' nu  $y_0$  ci  $y'_0$ , atunci efectul introducerii lui  $x$  va fi:

$$(y_n - y_0) - (y'_n - y'_0)$$

De asemenea, dacă pentru cazul (II) nu avem posibilitatea să variem efectiv variabila  $x$  în cadrul unui grup experimental, se pot alege mai multe grupuri experimentale, fiecare fiind caracterizat de o anumită stare a lui  $x$ .

În cazul experimentelor cu grup de control, principala dificultate ce trebuie învinsă este alegerea grupurilor de așa manieră încît să avem un grad suficient de ridicat de certitudine că, exceptînd acțiunea lui  $x$  în grupul experimental, toate celelalte transformări se vor realiza identic în ambele grupuri. De aceea se urmărește constituirea grupului de control, astfel încît să reflecte cît mai fidel structura grupului experimental, să se ajungă la o cît mai perfectă similitudine a lor. Există, practic, două procedee principale pentru atingerea acestui deziderat:

- procedeul potrivirii pe perechi și

- procedeul probabilistic

În primul caz, pentru fiecare individ din grupul experimental se alege un individ avînd caracteristici identice, în

x



grupul de control (exemplu: același sex, vîrstă, pregătire școlară etc.) În al doilea, grupurile se aleg ca niște egantioane aleatoare din aceeași populație, mizîndu-se pe faptul că întîmplarea va potrivi suficient de bine lucrurile, astfel încît să se cuprindă în aceleași proporții anumite tipuri de indivizi în fiecare grup.

Se observă că riscurile sînt mari în ambele situații. În cazul potrivirii pe perechi, putem controla doar un număr relativ restrîns de factori (aceia care efectiv sînt manifesti) și deci putem omite unii foarte importanți. În al doilea, așa cum am sugerat deja, ne încredem într-o șansă care ar putea să ne joace cele mai neplăcute feste. Iar probabilitatea să se întîmple așa este cu atît mai mare cu cît grupurile sînt mai reduse ca dimensiune. Or, în cele mai frecvente cazuri, nu avem posibilitatea de a experimenta decît tocmai cu asemenea grupuri restrînse.

N-am pomenit pînă aici nimic despre factorul C. Este evident că dacă acțiunea sa este considerată a fi semnificativă, e nevoie ca această acțiune să se manifeste cu egală intensitate în grupul experimental și în cel de control. Adică e bine, dacă cei din G știu să participe la un experiment, s-o știe și cei din G' și să reacționeze la fel. Este tocmai ideea subliniată de A. Miha (1973), care sugerează administrarea "unui pseudo (sau surrogat) x și în grupul de control" (p.105)

Mai menționez, în sfîrșit, și dificultățile legate de cea de a treia condiție pe care am subliniat-o pentru experiment: caracterul științific. Foarte multe obstacole sînt ridicate de gradul scăzut de dezvoltare a teoriei sociologice, ceea ce are repercusiuni asupra alegerii fenomenului de studiat, asupra consistenței ipotezelor etc. Alte dificultăți sînt legate de scopul cercetării. Experimentarea în societate presupune aproape întotdeauna o dependență a cercetărilor de factorii de decizie, deci subordonația interesului științific celui practic, cu toate conștințele ce pot uneori decurge de aici asupra rigorii științifice.



#### 4. Experimentul de laborator și experimentul de teren

Literatura metodologică a științelor sociale face, de regulă, o distincție între aceste două tipuri de experiment. În ceea ce mă privește, țin să-o precizez de la bun început, nu mi se pare că o atare diferențiere s-ar cere cu atîta putere subliniată, datorită faptului că în practică delimitarea nu apare totdeauna prea netă. Dificultățile în distingerea celor două tipuri este legată de alegerea unui criteriu ferm de departajare. De regulă, este invocat unul sau altul dintre următorii factori, ținînd cont de o dihotomizare a stării fiecăruia:

- (i) Mediul: artificial sau natural
- (ii) Izolarea: puternică sau slabă
- (iii) Situația: inedită sau naturală
- (iv) Factorul C: prezent sau absent

Se vorbește deci de experiment de laborator, în contrast cu cel natural, atunci cînd subiecții sînt: urmăriți într-un mediu artificial și/sau puternic izolați de influențe exterioare și/sau supuși unor situații inedite și/sau conștienți de faptul că sînt supuși experimentului.

Este clar că aceste dihotomii sînt prea rigide pentru a se putea identifica întocmai în practică. Faptul că, de exemplu, în cadrul unei întreprinderi, luăm un grup de indivizi și-l plasăm într-o începere separată înseamnă cu siguranță alterarea mediului lor natural de muncă, dar nu este vorba nici de condiții cu totul artificiale. Cît privește ultima cerință, astăzi se recomandă, din mai multe motive, dar mai ales de ordin etic, ca orice experiment să se facă numai cu acordul subiecților, deci, practic, ultimul criteriu este din ce în ce mai puțin operant. Probabil că cea mai evidentă situație în care se poate vorbi de experiment de laborator este cea indicată de factorul (iii), adică atunci cînd subiecții sînt supuși să joace anumite roluri pe care ei nu le îndeplinesc în viața de zi cu zi, căci aceasta implică în fapt și celelalte cerințe.

Chiar dacă se acceptă distincția dintre cele două genuri



de experiment, trebuie precizat că atunci când se vorbește de experimentul de laborator nu e neapărat nevoie să ne imaginăm că indivizii sînt plasați într-o sală pe a cărei ușa scrie "Laborator" că sînt înconjuțați de tot felul de aparate etc., deși există evident și astfel de laboratoare. Ele sînt descrise pe larg în literatura de specialitate - a se vedea, de exemplu, S.Chelcea (1982) - și nu voi insista asupra modului lor de dotare și organizare.

Ceea ce aș vrea să subliniez aici este o problemă epistemologică importantă, legată de experiment și exprimată adesea în discuțiile despre valoarea și oportunitatea experimentelor de laborator. Se reproșează frecvent acestei din urmă forme de experimentare faptul că ar scoate indivizii din situația naturală, plasîndu-i în una artificială, ceea ce ar altera raporturile dintre ei și s-ar ajunge la concluzii fără nici o aplicabilitate. Iată ce spune în acest sens T.Nicola (1973), care, după ce includea printre elementele ce constituie "esența experimentului științific" starea pură și crearea artificială de condiții (p.7), face o cotitură de 180 de grade, susținînd că experimentul de laborator "este limitat prin capacitatea sa de a surprinde acea autenticitate a realului ....., a sesiza și desprinde esența fenomenului social și, prin urmare, este limitat în posibilitățile de a generaliza rezultatele dobîndite, de a fundamenta și construi noi teorii" (p.23). Aș mai aminti că, în aceeași lucrare, autorul nega posibilitatea observației de a ajunge la esența fenomenelor, punînd superioritatea metodei experimentale tocmai pe asemenea studii în stare "pură" (p.4).

Avem de a face în toate criticile de genul celei de mai sus (chiar dacă sînt perfect consecvente) cu o neînțelegere a ceea ce înseamnă o legitate a naturii (socialului). Legea, ca element ce ține de esență - o știm cel puțin de la Hegel încoace - nu se manifestă în mod vizibil constant în fenomene, acestea trebuind decorticate pentru a ajunge la esență, Nici pentru fizică nu se întîlnesc în natură condițiile de laborator, dar pentru asta nimănui nu i-a trecut prin cap să critice experimentul de laborator pentru că nu reproduce condițiile reale, nimeni nu respinge, de exemplu, legea căderii corpurilor în vid pentru motivul că pe pămînt nu există vid și deci nicăieri fulgul nu va ajunge la sol odată cu piatra, cînd le dăm drumul să cadă în același moment și



de la aceeași înălțime. A critica experimentul că nu surprinde "autenticitatea realului" înseamnă a nu sesiza esența acestei metode de cercetare științifică.

Leon Festinger(1963) argumentează și el în același sens. Experimentul de laborator "nu are și nu poate avea pretenția de a reproduce o situație a vieții reale. Dacă vrem să studiem un aspect al unei situații realmente trăite, ar fi o nebunie să vrem să aranjăm într-un raport condiții copiate de pe această situație. De ce să nu se meargă direct la situație și să se studieze așa cum este ea însăși. Experimentul de laborator vizează crearea unei situații în care operarea cu anumite variabile va putea fi pusă în lumină datorită unor condiții speciale, bine identificate și definite în mod clar. Puțin importantă că o astfel de situație nu se întâlnește în viața reală" (p.166).

Dacă ar fi să vorbim despre limitele așa-ziselor experimente de laborator, atunci ar trebui să subliniem că eventuala lor irelevanță se datorează tocmai dificultăților de a întruni în mod riguros condițiile experimentării și, în speță, cele referitoare la izolarea și controlul factorilor. De asemenea, este evident că doar o parte din fenomenele sociale se pretează la astfel de experimentări, întrucât este greu de văzut cum pot fi îndeplinite cerințele experimentului în cazul unor fenomene ce se manifestă numai prin antrenarea unor mase mari de oameni.

De aici rezultă și concluzia generală privind posibilitatea, limitele și oportunitatea experimentării în sociologie. Dacă se vor inventaria cazurile de experimente indicate în literatura științelor sociale (ceea ce n-am mai făcut aici din lipsă de spațiu și pentru a nu repeta lucruri pe care oricine le poate citi în lucrările indicate la bibliografie) se va ajunge la următoarele constatări:

a) În cele mai numeroase cazuri, studiile experimentale respective aparțin unor domenii învecinate sociologiei, ca psihologia socială (mai cu seamă problematicii psihosociologice a grupurilor mici), ergonomia, știința conducerii etc.

b) Așa-numitele experimente sociale de mare anvergură sînt, aproape în totalitatea cazurilor, de fapt pseudoexperimente, deoarece nu întrunesc în chip satisfăcător condițiile de bază ale a-



cestei metode științifice.

De aici nu rezultă o respingere de plano a experimentului ca metodă de cunoaștere sociologică, ci doar sublinierea dificultăților imense ce stau în fața sociologului care ar intenționa să uzeze de acest gen de investigație. De aici <sup>nu</sup> rezultă, de asemenea, că trebuie să punem în întregime relativă întârziere în dezvoltarea sociologiei pe seama dificultăților în realizarea practică a metodei experimentale. Există, așa cum am mai spus, științe ale naturii care nu utilizează experimentul, ceea ce nu le-a împiedicat să aibă o istorie multimilenară și să ajungă la rezultate remarcabile.

Probabil că sociologii și alți specialiști din științele sociale consideră că atribuirea denumirii de experiment unor serii de tehnici de cercetare care nu au nimic de a face cu metode experimentale ar avea darul să ridice prestigiul investigațiilor lor. Este aceasta însă o tactică ce, pînă la urmă, se poate întoarce împotriva celui care o utilizează și a sociologiei în ansamblul său.



### PARTEA A III-A

#### ANALIZA STATISTICA A DATELOR

##### Cap. VII. CARACTERISTICI

Noțiunea de "caracteristică", încetățenită de multă vreme în limbajul statisticii descriptive, nu reprezintă altceva decât trăsătura, însușirea etc. în funcție de care s-a realizat o clasificare sau o măsurare propriu-zisă în cadrul unei mulțimi de indivizi ce constituie eșantionul sau populația investigată de cercetător.

Pe baza unor considerații de ordin foarte practic - în sensul că ele provin exclusiv din modalitățile tehnice de lucru cu datele -, în statistică se face o distincție netă între caracteristicile calitative și cele cantitative, primele corespunzând acelor trăsături care s-au abordat cu ajutorul scalelor nominală și ordinală, iar ultimele celor ce corespund scalelor de intervale și de rapoarte. În esență, deosebirea aceasta se datorează imposibilității sau posibilității de a utiliza valorile numerice, atribuite treptelor scalei, în operații matematice obișnuite. Chiar dacă, așa cum am arătat, există diferențe mari între scalele nominale și cele ordinale sau între cele de intervale și de rapoarte, chiar dacă, pe de altă parte, scala ordinală presupune deja manifestarea cantitativă a trăsăturii studiate, apropiindu-se deci de cea de intervale, în operațiile statistice distincția menționată între cele două tipuri de caracteristici este foarte tranșantă, traducându-se - cum vom vedea imediat - în procedee diferite de tratare a datelor.

În scopul simplificării limbajului, voi folosi, în cele ce urmează, pentru caracteristicile cantitative termenul, de asemenea încetățenit, de variabilă statistică, sau, pur și simplu, de variabilă.



## 1. Distribuții de frecvență

Aplicarea unei scale nominale sau ordinale la o colectivitate statistică ne conduce imediat la o clasificare a indivizilor, adică la o repartizare a acestora pe treptele scalei respective. Spre deosebire de operațiile de clasificare folosite în general în diferite științe, clasificările de tip statistic implică și precizarea cuantumului de indivizi ce se plasează pe diferite trepte ale scalei. Prin urmare, noțiunea de caracteristică implică, așa cum am subliniat de la început, nu numai o trăsătură sau, eventual, un pachet de trăsături după care se efectuează clasificarea ci și modalitatea în care indivizii se distribuie în clasele respective. Tocmai acest aspect cantitativ, legat de numărarea indivizilor din diferite clase și de operațiile<sup>cu</sup> numerele astfel obținute, constituie conținutul tuturor procedurilor statistice referitoare la caracteristicile calitative.

În modul cel mai general vorbind, construcția unei caracteristici calitative,  $A$ , înseamnă realizarea unei corespondențe între o mulțime:

$$\mathcal{A} = \{A_1, A_2, \dots, A_s\}$$

constituită din clasele  $A_1, A_2, \dots, A_s$  ale scalei și o mulțime:

$$\mathcal{K} = \{k_1, k_2, \dots, k_s\}$$

de numere naturale, ale cărei elemente arată câți dintre indivizii populației studiate se plasează pe fiecare treaptă a scalei. Mulțimea  $\mathcal{K}$ , ale cărei elemente se numesc frecvențe sau frecvențe absolute, reprezintă distribuția de frecvență a caracteristicii  $A$ .

O asemenea corespondență se scrie, de regulă, sub forma unui tabel ca cel de mai jos:

Clasele	$A_1$	$A_2$	...	$A_s$	Total
Frecvențele	$k_1$	$k_2$	...	$k_s$	$n$



Pe prima linie se trec simbolurile claselor constituite iar pe a doua frecvențele,  $k_1$ , a căror sumă trebuie să fie egală cu  $n$ , volumul populației studiate:

$$k_1 + k_2 + \dots + k_s = n$$

sau, prescurtat

$$\sum_{i=1}^s k_i = n$$

Relația de mai sus provine evident din principiul clasificării, conform căruia fiecare individ din populație trebuie să se plaseze în una și numai în una din clase.

Forma cea mai simplă a unei caracteristici calitative este cea dihotomică, realizată atunci când populația se divide numai în două clase. De obicei, o astfel de caracteristică poartă numele de atributivă sau de atribut, deoarece aproape întotdeauna se poate considera că indivizii plasați în una din clase posedă un atribut, pe când ceilalți nu-l posedă. În acest caz, se poate folosi pentru clase notația  $A$  (pentru cei care au calitatea respectivă) și  $\bar{A}$  (pentru cei care n-o au). Tabelul devine atunci:

Clasele	$A(A_1)$	$A(A_2)$	Total
Frecvențele	$k_1$	$k_2$	$n$

Pentru variabilele statistice situația este puțin diferită. Elementele scalei sînt numere care se obțin printr-o operație de măsurare propriu-zisă sau prin alte proceduri de atribuire. Aceste valori ale scalei pot fi în număr finit sau infini, plasează pe un interval ce merge de la valoarea cea mai mică la cea mai mare. Când o variabilă acoperă, potențial vorbind, cu valorile sale un întreg interval  $[a, b]$ , ea se numește continuu, iar atunci când ia numai anumite valori, de regulă în număr finit, discreet.

Este evident că în acțiunile concrete de cercetare nu vom întâlni decît variabile discrete, deoarece mijloacele noastre de investigație și percepția umană nu pot surprinde o infinitate (con-



tinuă) de nuanțe. Aceasta nu înseamnă că trebuie să renunțăm la distincția dintre cele două tipuri de variabile. De pildă, vârsta, măsurată pe axa timpului, este clar o variabilă continuă, ce poate lua valori de la 0 la 100 și ceva de ani, deși practic este imposibil să precizăm vârsta cuiva cu o exactitate de ordinul secundelor. Dimpotrivă, dimensiunea unei gospodării este o variabilă discretă clasică, singurele valori posibile fiind: 1, 2, 3 etc. Chiar și unele variabile discrete, prin definiție, dar cu un număr foarte mare de valori pot fi asimilate celor continue. De exemplu, retribuirea, presupunând că nu se calculează în subdiviziuni monetare mai mici de 1 leu, atinge o mulțime extrem de mare de valori, de ordinul miilor, și ca atare poate fi practic considerată o variabilă continuă, dacă interesele studiului cer acest lucru.

În cazul variabilelor discrete cu un număr relativ mic de valori, acestea din urmă (valorile) se substituie, formal, simbolurilor claselor și situația poate fi prezentată ca și pentru caracteristicile calitative:

Valori	$x_1$	$x_2$	...	$x_s$	Total
Frecvențe	$k_1$	$k_2$	...	$k_s$	$n$

Dacă variabila este continuă sau cu un număr foarte mare de valori, sistematizarea datelor devine practic imposibilă dacă în tabel s-ar trece valoarea exactă a fiecărui individ în parte. În astfel de cazuri, se apelează la un artificiu, și anume se construiesc clase prin gruparea tuturor valorilor dintr-un subinterval; se obțin astfel caracteristici cu intervale de grupare. Este de notat că aceste caracteristici își pierd într-o oarecare măsură atributul de variabile, întrucât fiecare interval-clasă nu mai este caracterizat de o valoare numerică așa cum cerea definiția, ci de o mulțime de valori. Intervalele corespund mai degrabă unor calități, avînd în vedere că toți indivizii plasați într-un același interval sînt considerați identici, din punctul de vedere al caracteristicii studiate. Dar cum trecerea de la variabile la caracteristici calitative înseamnă de fapt un regres (pentru că primele se pretează la infinit mai multe operații și analize statistice), atunci cînd cercetarea nu se oprește la simpla descriere



a caracteristicilor, se procedează la o revenire spre cantitativ, atribuindu-i-se fiecărui interval o singură valoare numerică, de regulă una centrală intervalului.

Prin urmare, trecerea de la variabile propriu-zise la caracteristici cu interval de grupare și revenirea la variabile presupun două operații care se expun unui grad mai mare sau mai mic, dar totdeauna nonneglijabil, de convenționalitate: alegerea limitelor intervalelor și a valorilor caracteristice pentru acestea. La prima operație intervin o serie de elemente de apreciere ce pot deriva din teoria asupra fenomenului analizat și/sau din distribuția empirică a indivizilor după variabila originară. Astfel, clasele reprezentate de intervale pot fi impuse de considerente ținând de identitatea calitativă a indivizilor, care, chiar dacă iau valori diferite, pot fi considerați de aceeași natură. De pildă, dacă variabila este "vîrsta" și o utilizăm într-o cercetare ce vizează aspecte legate de activitățile școlare, putem forma intervale de vîrstă care să corespundă preșcolarilor, elevilor din învățămîntul primar, gimnazial, și liceal, studenților etc. Dacă nu avem nici un motiv de asemenea natură care să sugereze grupările, atunci vom folosi asemenea intervale (nu neapărat de lungime egală) încît să se cuprindă în fiecare un număr suficient de mare de cazuri pentru a putea continua analizele.

Există și unele formule - empiric stabilite - care ne pot ajuta în evaluarea lungimii optime a intervalelor, cum este cea a lui Sturges:

$$i = \frac{x_{\max} - x_{\min}}{1 + 3,322 \cdot \lg n}$$

unde  $i$  este lungimea intervalului,  $x_{\max}$  și  $x_{\min}$ , cea mai mare și respectiv cea mai mică valoare luată în populație iar  $\lg n$  logaritmul zecimal din numărul total de indivizi. Valoarea unor asemenea formule este destul de limitată; uneori ele pot fi folosite cu folos, alteori nu. De exemplu, dacă variabila  $x$  este vîrsta, cu valori între 0 și 100 ani, pentru un eșantion de 1000 persoane ( $\lg 1000 = 3$ ) vom avea:



$$i = \frac{100}{1 + 3,322 \cdot 3} \approx \frac{100}{11} \approx 9$$

Pentru comoditate, se poate lua intervalul de 10 ani și atunci se obține o clasificare acceptabilă. În schimb, dacă  $x$  este retribuția, cu valori între, să zicem, 1700 și 5000 lei, pentru același esantion de 1000 persoane obținem  $i = 3300 : 11 = 300$  lei, deci 33 de intervale, ceea ce este foarte mult pentru o analiză obișnuită.

Formula lui Sturges presupune că intervalele de grupare sînt egale, lucrul normal pentru că tocmai astfel de situații sînt cel mai adesea întîlnite. Dar nu se întîmplă mereu așa. Uneori indivizii sînt concentrați foarte puternic într-o anumită zonă a scalei și se găsesc în număr mic pe alte părți întinse ale ei. Se cer formate atunci intervale de lungimi inegale, împărțind în secțiuni mai mici zona populată și mai mari cea slab populată. În fine, o ultimă observație referitoare la intervale privește grija pe care trebuie să o avem atunci cînd fixăm capetele acestora, pentru a putea plasa exact indivizii careiau valori în zona de întîlnire a două intervale. Singurele excepții de la această regulă o constituie uneori primul și ultimul interval, care pot rămîne deschise la stînga și, respectiv, la dreapta, dacă nu se poate preciza valoarea minimă și maximă.

Revenind la distribuțiile de frecvență, trebuie menționat că cele mai elementare calcule ce se pot face cu ele - indiferent de natura caracteristicii - constau în determinarea frecvențelor relative. Frecvența relativă a unei clase a caracteristicii este fracțiunea (ponderea) indivizilor din populația totală care aparțin acestei clase. Dacă notăm cu  $p_i$  frecvențele relative, atunci:

$$p_i = \frac{k_i}{n} \quad (i = 1, 2, \dots, s)$$

adică frecvența relativă se obține prin împărțirea frecvenței absolute la numărul total de indivizi din populație.

Rezultă imediat două proprietăți importante ale frecvențelor relative:



a) ele sînt valori pozitive subunitare ( $0 < p < 1$ );

b) suma lor este egală cu unitatea:  $p_1 + p_2 + \dots + p_s = 1$

Tot o speță de frecvență relative sînt și procentele. În loc de a se reduce totalul la unitate, ca în cazul precedent, pentru procente el se reduce la 100. În consecință, procentul de indivizi dintr-o clasă se va calcula:

$$\text{procentul} = \frac{k_1}{n} \cdot 100 = p_1 \cdot 100$$

Frecvențele relative și procentele se utilizează pentru două scopuri fundamentale:

a) Să sesizăm mai ușor care este ponderea diferitelor clase în populație. Astfel, dacă într-o clasă avem 348 indivizi, dintr-o populație de 4350, prin calcularea procentului (8%) vom înțelege mai ușor gradul de cuprindere al acestei clase.

b) Mai folositoare sînt frecvențele relative și procentele atunci cînd dorim să comparăm caracteristici identice construite asupra unor populații de mărimi diferite. Comparația este practic imposibilă cu frecvențe absolute deoarece aceste mărimi au semnificații diferite în funcție de total. Calculul frecvențelor relative și al procentelor constituie deci o procedură de standardizare a informației dintr-un tabel statistic. Pe această cale se poate constata că cei 348 de indivizi de mai înainte reprezintă aceeași fracțiune, în populația de 4350, ca și alți 64 indivizi dintr-o populație de 800.

Frecvențele relative (la unitate) se folosesc mai ales ca mărimi intermediare în alte operații de analiză statistică la care sînt supuse caracteristicile sau ca valori interpretabile în termeni de probabilități. Procentele își găsesc cea mai curentă utilizare în prezentarea datelor în rapoarte de cercetare, articole, cărți etc., deoarece ele sînt mai ușor de înțeles de către publicul larg. În adevăr, pentru cei mai mulți oameni este mai ușor să înțeleagă semnificația aserțiunii "în anul cutare, în țara cutare, în medie din 100 de copii născuți, 51 au fost băieți și 49 fete (sau 51% dintre născuți au fost băieți și 49% fete)" decît a propoziției: "în anul cutare, în țara cutare, ponderea băieților în rîndul nou-născuților a fost de 0,51 iar a fetelor de 0,49", deși



informația transmisă sub cele două învelișuri formale este absolut identică.

În continuare, voi trece la prezentarea unor exemple concrete, în scopul ilustrării celor spuse până aici. Fiind, vorba, în esență, de o ilustrare prin intermediul tabeleleor, îmi permit pentru a nu mai reveni cu altă ocazie, să precizez acum câteva reguli elementare cu privire la utilizarea și prezentarea tabelelor în cadrul lucrărilor științifice:

- Fiecare tabel primește un număr. Numerotarea poate merge în continuare de la începutul până la sfârșitul lucrării sau poate fi reluată de la 1 în cadrul fiecărui capitol sau secțiuni. În acest din urmă caz, este recomandabil ca, pe lângă numărul de ordine al tabelului, să se indice și cel al capitolului, pentru a se facilita trimiterile în text la tabele. De exemplu, scriind "Tabelul 2.7", se va înțelege că este vorba de tabelul 7 din capitolul 2.

- Fiecare tabel are un titlu în care se precizează caracteristica descrisă, populația asupra căreia a fost calculată, tipul de frecvențe sau alți indicatori folosiți în tabel. Titlul tabelului urmează după numărul de ordine al tabelului, aceste informații putând fi scrise deasupra sau dedesubtul său, mai aproape de tabel decât de restul textului, pentru a se putea citi ușor.

- Dacă tabelul conține date ce nu sînt culese de către autor, se va specifica sursa de proveniență a acestora.

- Capetele liniilor și coloanele tabelului este bine să conțină o informație care, coroborată cu cea din titlu, să permită cititorului să înțeleagă tabelul fără a face apel la textul lucrării.

Să începem cu caracteristicile calitative și mai exact cu cazul dihotomic. Tabelele 1 și 2 reprezintă repartitia populației R.S.R., în 1982, după caracteristicile "sex" și "mediu", iar tabelul 3 răspunsurile - de forma Da-Nu- la o întrebare de opinie dintr-un chestionar.



Sexul	Bărbați	Femei	Total
Precvențe	11.091.465	11.386.238	22.477.703
%	49,3	50,7	100

Tabelul 1. Populația R.S.R. pe sexe în 1982 Sursa: Anuarul Statistic al R.S.R. 1983

Mediul	Urban	Rural	Total
Precvențe	11.575.632	10.902.071	22.477.703
%	51,5	48,5	100

Tabelul 2. Populația R.S.R. pe medii în 1982. Sursa: Anuarul statistic al R.S.R. 1983

Răspunsul	DA	NU	Total
Precvențe	900	300	1200
%	75,0	25,0	100

Tabelul 3. Răspunsurile la o întrebare din chestionar  
(date fictive)

Pentru caracteristicile calitative cu mai multe clase, avem, de regulă, două mari situații: când scala este nominală și atunci clasificarea apare ca o structurare orizontală a populației și când scala e ordinală, ceea ce implică o structurare pe verticală (sau pe nivele ierarhice) a acesteia. Un exemplu din prima categorie este dat în tabelul 4, iar a doua în tabelul 5.



Starea civilă	Necăsătorit	Văduv	Divorțat	Total
Precvențe	149.103	5.261	20.083	174.447
%	85,5	3,0	11,5	100

Tabelul 4. Populația masculină din R.S.R. care s-a căsătorit în 1982, după starea civilă înainte de căsătorie. Sursa: Anuarul Statistic al R.S.R. 1983

Răspunsuri	Foarte puțin	Puțin	Moderat	Mult	Foarte mult	Total
Precvențe	100	400	500	400	200	1600
%	6,25	25,00	31,25	25,00	12,50	100

Tabelul 5. Distribuția răspunsurilor la o întrebare din chestionar. (date fictive)

Următoarele tabele prezintă trei cazuri de caracteristici cantitative. În tabelul 6 apare o caracteristică discretă pentru care sînt indicate exact toate valorile. Tabelele 7 și 8 cuprind două exemple de caracteristici cu intervale de grupare, ultimul prezentînd și particularitatea că intervalele de la margine sînt deschise.

Nr.de persoane în gospodărie	1	2	3	4	5	6	7	8	9	Total
Nr.de gospodării	30	40	60	50	30	20	10	5	5	250
%	12,0	16,0	24,0	20,0	12,0	8,0	4,0	2,0	2,0	100

Tabelul 6. Distribuția gospodăriilor unei localități rurale după numărul membrilor ce le compun(date fictive)



Medii	2 - 2,99	3 - 3,99	4 - 4,99	5 - 5,99	6 - 6,99	7 - 7,99	8 - 8,99	9 - 10	Tot.
Nr. de candidați	30	40	130	150	100	75	50	25	600
%	5,6	6,7	21,7	25,0	16,7	12,5	8,3	4,2	100,1

Tabelul 7. Distribuția mediilor obținute de candidații la admitere la facultate (date fictive)

Sub 15 Vîrsta ani	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50 și peste	To- tal
Frec- vențe	638	46.693	146.103	95.939	38.621	12.324	3.720	302	29 344.369
%	0,2	13,6	42,2	27,8	11,2	3,6	1,1	0,1	100

Tabelul 8. Distribuția născuților vii în R.S.R., în 1982, după grupa de vîrstă a mamei. Sursa: Anuarul Statistic al RSR. 1983

## 2. Erafece atașate tabelelor de frecvențe

Forma fundamentală de prezentare a informației obținute în legătură cu o caracteristică este tabelul de frecvențe. Pentru o ilustrare mai sugestivă a mărimii frecvențelor, a raporturilor și diferențelor de populare a claselor, a modului cum se schimbă frecvențele cînd variabila trece de la valorile mici spre cele mari etc., tabelelor de frecvențe li se pot atașa diferite forme grafice<sup>x)</sup> Există foarte multe procedee de construcție a graficelor, chiar pentru o aceeași distribuție de frecvențe, dar acestea, în cea mai mare parte a lor, nu au decît o importanță relativ minoră, pur ilustrativă.

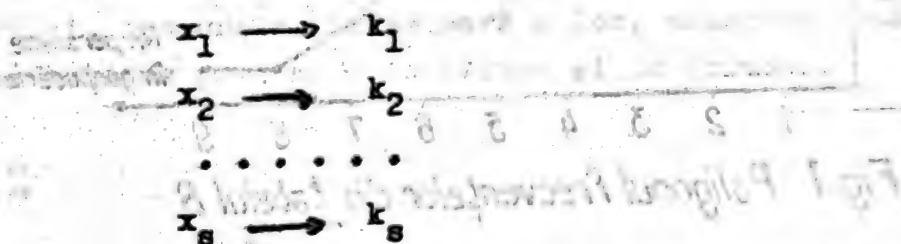
În cadrul rapoartelor de cercetare, însoțirea tabelelor de forme grafice este aproape obligatorie, deoarece beneficiarul poate astfel să înțeleagă mult mai ușor și mai rapid esența tabelului. În studiile destinate publicațiilor științifice, în lucrările de diplomă sau în tezele de doctorat, folosirea figurilor grafice este stricto sensu și tabelul este o formă de reprezentare grafică a



fice este recomandată doar în cazul unor tabele mai complexe, care presupun un efort mai mare spre a fi citite sau în situația unor variații interesante în rândul frecvențelor.

În cele ce urmează voi prezenta doar câteva forme grafice principale, deoarece în legătură cu această problemă există o literatură destul de bogată. (A se vedea, de exemplu, Al. Bărbat, 1972). Forma fundamentală de prezentare grafică a datelor pentru o variabilă statistică este poligonul frecvențelor.

Așa cum am văzut la început, orice caracteristică este o funcție de genul  $f: A \rightarrow K$ , ce aplică deci mulțimea claselor în mulțimea numerelor naturale. În cazul variabilelor statistice, mulțimea  $A$  devine o submulțime de numere reale și deci o caracteristică de acest gen corespunde obişnuitelor funcții studiate în școala generală și liceu. Într-adevăr, un tabel pentru o variabilă statistică  $X$  este o corespondență de forma:



deci o mulțime de perechi ordonate  $(x_i, k_i)$ , care se pot ușor reprezenta sub formă grafică.

Dacă luăm un sistem de axe de coordonate ortogonale, în care pe axa absciselor vom reprezenta valorile variabilei iar pe cea a ordonatelor frecvențele, atunci mulțimea perechilor ordonate de mai sus primește drept imagine o mulțime de puncte în plan cu abscisele  $x_i$  și ordonatele  $k_i$ . Pentru a se sugera trecerea de la o valoare  $x_i$  la valoarea următoare  $x_{i+1}$ , se unesc succesiv, două câte două, aceste puncte cu segmente de dreaptă, obținându-se o linie poligonală ce reprezintă tocmai forma grafică menționată: poligonul frecvențelor.

De exemplu, distribuției din tabelul 6,1 se va atașa graficul din figura<sup>1</sup>. Dacă vom avea de a face cu intervale de grupare,

(continuare)  
datelor. Cu toate acestea, noțiunile de grafic sau de reprezentare grafică s-au încetățenit în limbajul curent de așa manieră încât nu mai includ "tabelele" sau "prezentarea tabelară" a datelor.



pe axa absciselor se vor trece aceste intervale, iar verticalele vor fi ridicate din punctele de origine alese. Astfel, pentru distribuția din tabelul 7 se poate construi un grafic ca cel din figura 2.

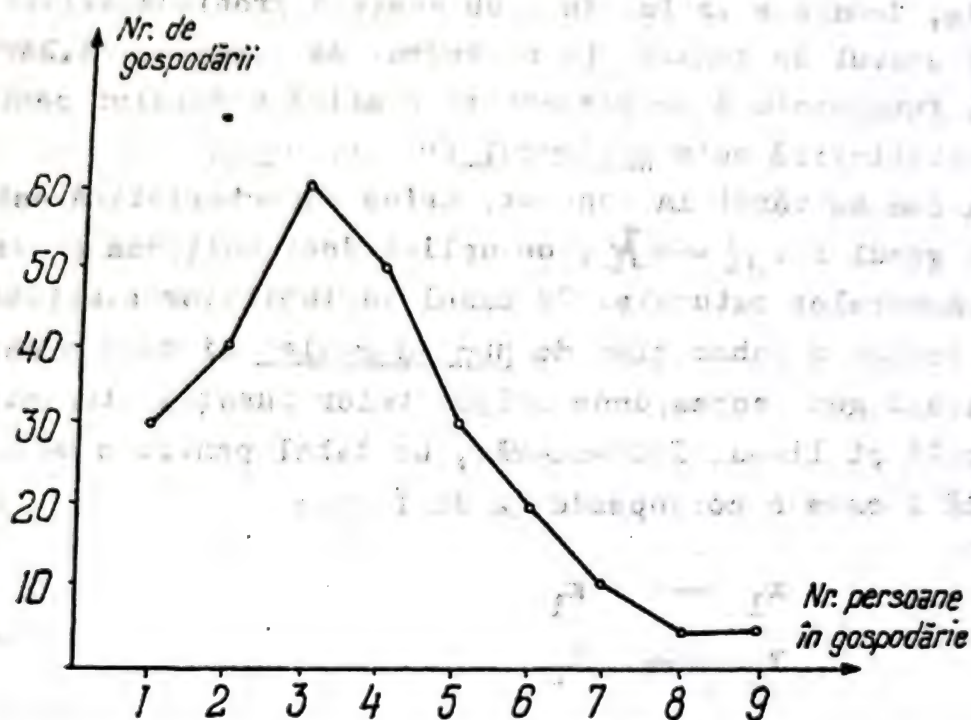


Fig. 1 Poligonul Frecvențelor din tabelul 6

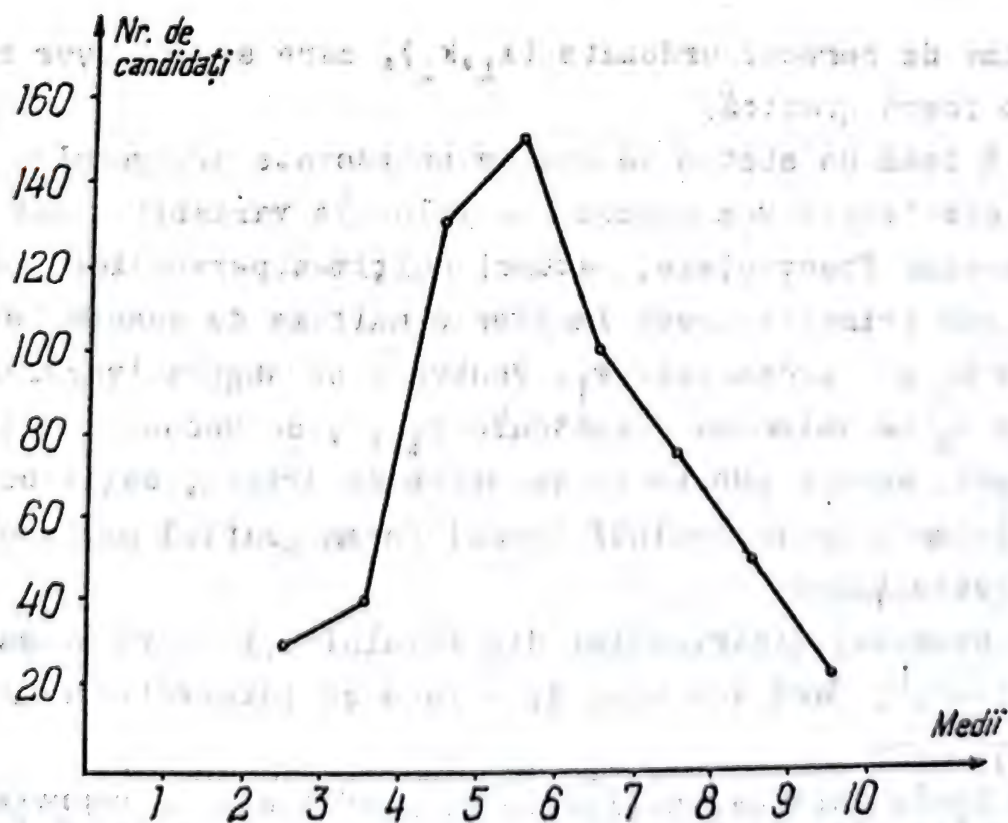


Fig. 2 Poligonul Frecvențelor din tabelul 7



Poligonul frecvențelor - ca de altfel toate formele grafice atașate distribuțiilor de frecvență - se poate construi fie utilizând frecvențele absolute, cum am procedat în fig.1 și 2, fie cele relative, fie procentele. Acestea din urmă se recomandă mai ales acolo unde frecvențele absolute prezintă variații foarte mari în rubricile tabelului. Dar utilizarea frecvențelor relative și procentelor devine indispensabilă dacă dorim să ilustrăm diferența de "comportament" dintre două populații diferite, în raport cu o aceeași variabilă, în astfel de cazuri, indiferent că vom trasa cele două linii poligonale pe aceeași sistem de axe sau vom folosi două grafice diferite, comparațiile nu sînt posibile - așa cum am mai spus - decît în termenii valorilor relative.

Să reprezentăm grafic și datele tabelului 8, cu ajutorul procentelor (fig.3). Se observă că punctele de origine pentru intervalele semideschise de la capete s-au ales foarte aproape de limita superioară și, respectiv, inferioară a lor, deoarece fenomenul urmărit nu se extinde prea mult în stînga și în dreapta.

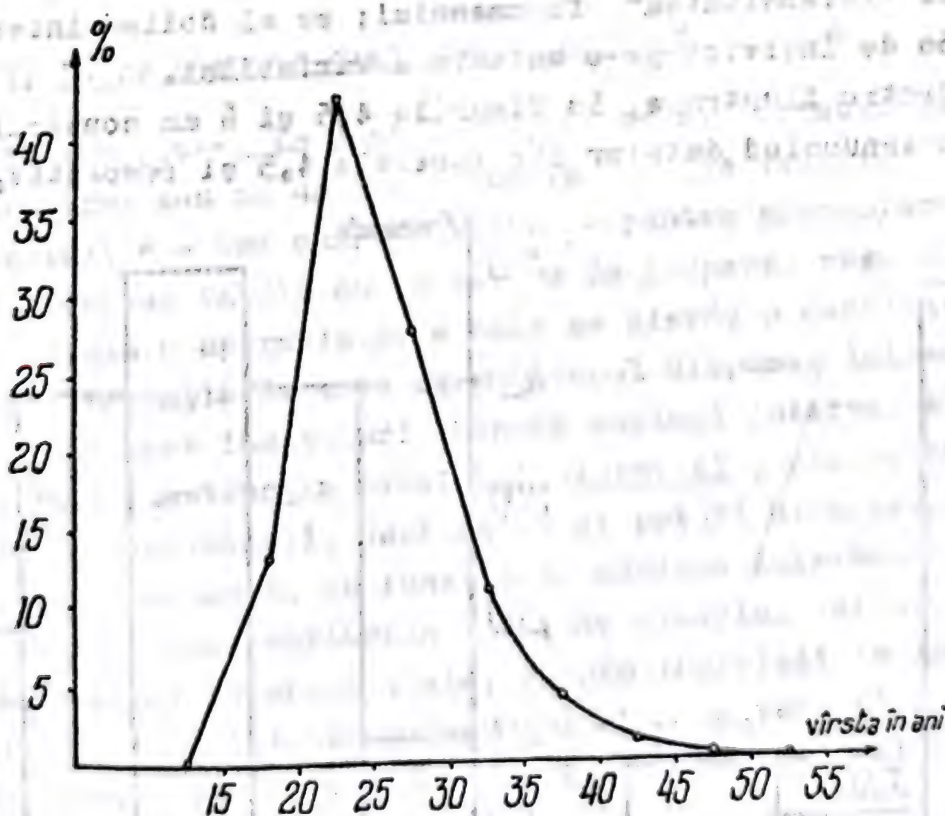


Fig. 3 Poligonul frecvențelor din tabelul 8



A doua formă grafică fundamentală pentru distribuțiile de frecvență este histograma. Ea se aplică, de regulă, caracteristicilor calitative și celor cantitative cu intervale de grupare când nu se alege un punct de origine pentru acestea. Principiul de construcție este și aici foarte simplu. Pe axa absciselor se reprezintă sub formă de segmente clasele caracteristicii sau intervalele de grupare, iar pe ordonată frecvențele (absolute, relative, procente). Se ridică pe fiecare interval de pe abscisă un dreptunghi cu înălțimea indicată pe ordonată de reprezentarea frecvențelor. Acest lucru este valabil, de fapt, numai când segmentele de pe abscisă sînt egale; în acest caz frecvențele sînt reprezentate atît prin înălțimea dreptunghiurilor cît și prin aria acestora. În cazul în care avem de a face cu intervale de grupare inegale, atunci bazele dreptunghiurilor variază în funcție de lungimea acestor segmente iar frecvențele pe intervale vor fi reprezentate numai de aria dreptunghiului. De pildă, dacă pe un interval de o unitate avem o frecvență de 100 indivizi și pe un altul de 2 unități <sup>aceeași</sup> frecvență, primul dreptunghi va avea înălțimea 100, pe cînd al doilea numai 50. Aici înălțimea dreptunghiului reprezintă "intensitatea" fenomenului; pe al doilea interval revin doar 50 de indivizi pe o unitate a variabilei.

Pentru ilustrare, în figurile 4, 5 și 6 am construit histograme corespunzînd datelor din tabelele 4, 5 și respectiv, 7.

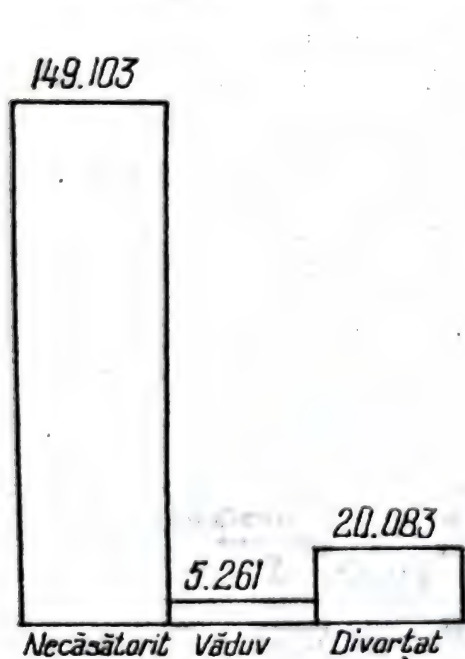


Fig. 4 Histograma frecvențelor din tabelul 4

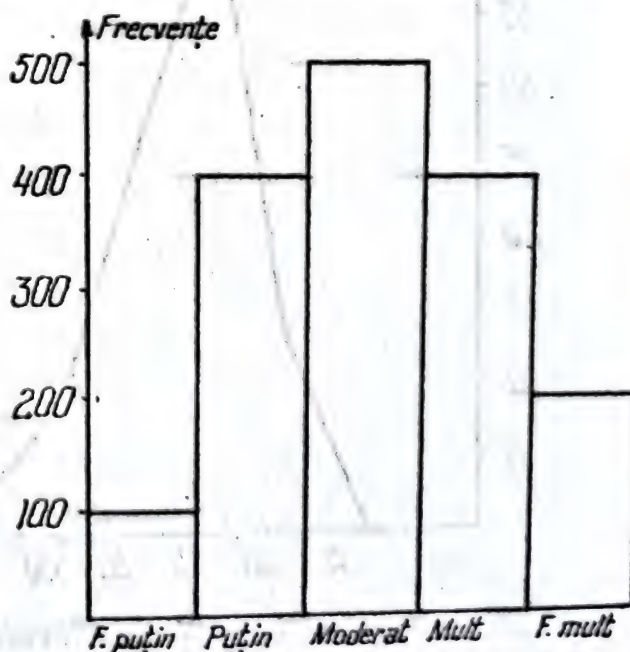


Fig. 5 Histograma frecvențelor din tabelul 5



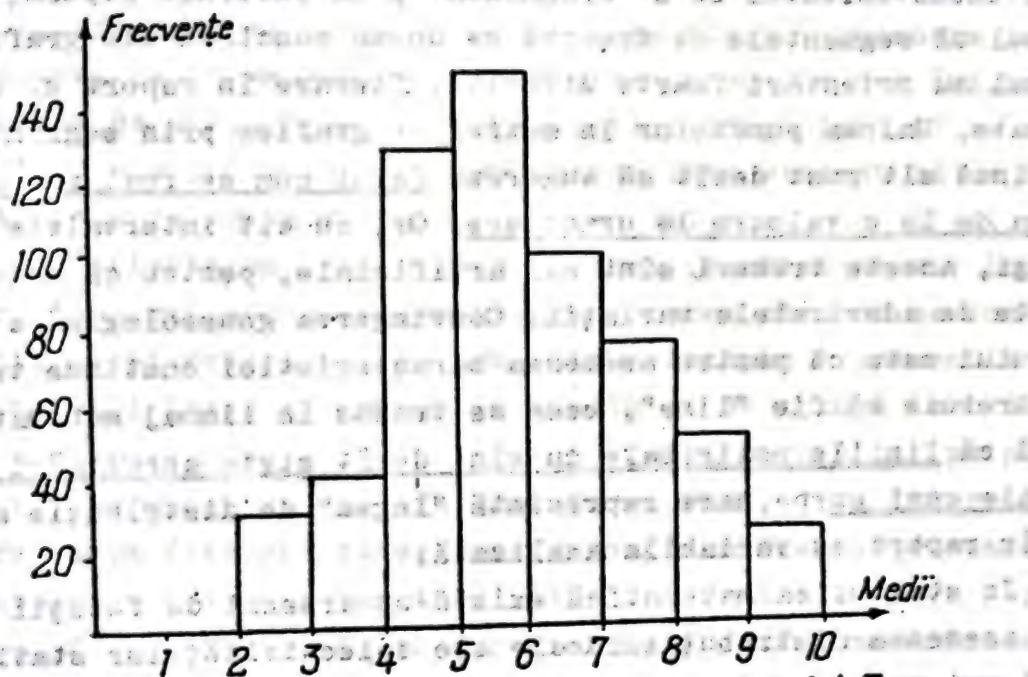


Fig. 6 Histograma Frecvențelor din tabelul 7

### 3. Curbe ataseate distribuțiilor de frecvență

Ne vom plasa, în continuare, în cazul variabilelor statistice continue sau cu un număr foarte mare de valori. Într-o asemenea situație — așa cum s-a văzut — pentru prezentarea datelor se construiesc tabele cu intervale de grupare, ceea ce, evident, constituie o operație prin care se pierde o cantitate însemnată din informația de care cercetătorul dispune. Într-adevăr, toți indivizii care iau valori într-un același interval sînt considerați, după construcția tabelului, identici, din punctul de vedere al caracteristicii, deși între ei pot fi diferențe mai mult sau mai puțin însemnate, în funcție de mărimea intervalului de grupare. Astfel, dacă variabila este, de exemplu, retribuirea și, pentru acoperirea întregii scale, grupăm indivizii în intervale de genul: 1500-1999 lei, 2000-2499 lei etc. se vede că doi indivizi sînt considerați ca avînd aceeași retribuire deși diferența între ei poate fi de aproape 500 lei.

Pe de altă parte, un număr redus de intervale de grupare — deși foarte favorabil pentru clasificări și analize ulterioare —



are și inconvenientul de a "fragmenta" prea puternic populația, în sensul că segmentele de dreaptă ce unesc punctele din graficul poligonal au orientări foarte diferite, fiecare în raport cu cele învecinate. Unirea punctelor în astfel de grafice prin segmente nu are însă alt rost decât să sugereze felul cum se realizează trecerea de la o valoare la următoarea. Or, cu cât intervalele sînt mai largi, aceste treceri sînt mai artificiale, pentru că mai îndepărtate de adevăratele variații. Convingerea gnoseologică a cercetătorului este că pentru asemenea caracteristici continue trecerile trebuie să fie "line", ceea ce tradus în limbaj matematic înseamnă că liniile poligonale nu sînt decît niște aproximații empirice ale unei curbe, care reprezintă "legea" de distribuție a populației în raport cu variabila analizată.

În statistica matematică există un arsenal de funcții care descriu asemenea distribuții ideale ale colectivităților statistice, după mai multe tipuri de curbe. Dintre acestea, cea mai cunoscută - și probabil cu aplicabilitatea cea mai largă - este așa-numita curbă a lui Gauss, sub formă de clopot, cunoscută și sub numele de curbă sau lege normală de distribuție. Două probleme importante se pun în legătură cu trecerea de la un set de date individuale (valori pentru indivizi empirici) la asemenea curbe matematice:

Mai întîi, una practică, și anume, avînd o distribuție empirică, se urmărește să se vadă dacă se poate afirma că această distribuție poate fi considerată o aproximație suficient de bună a uneia teoretică, mai exact trebuie testat dacă frecvențele empirice pot fi privite cu particularizări ale celor teoretice. Se consideră că lotul de indivizi studiați reprezintă un eșantion reprezentativ al unei populații, potențial infinite, iar testarea constă în calcularea semnificației diferențelor dintre frecvențele empirice și cele care ar rezulta, la același volum de populație și la aceleași intervale de valori, dacă distribuția urmează o curbă matematică ideală. Este deci o chestiune de testare a semnificației diferențelor, legată de eșantionare, despre care va fi vorba în ultima parte a cursului.

A doua problemă este una teoretică, de principiu, cerînd să se indice calea inversă, cea prin care s-ar putea trece de la valorile indivizilor empirici la o curbă. Presupunînd că popula-



are și inconvenientul de a "fragmenta" prea puternic populația, în sensul că segmentele de dreaptă ce unesc punctele din graficul poligonal au orientări foarte diferite, fiecare în raport cu cele învecinate. Unirea punctelor în astfel de grafice prin segmente nu are însă alt rost decât să sugereze felul cum se realizează trecerea de la o valoare la următoarea. Or, cu cât intervalele sînt mai largi, aceste treceri sînt mai artificiale, pentru că mai îndepărtate de adevăratele variații. Convingerea gnoseologică a cercetătorului este că pentru asemenea caracteristici continue trecerile trebuie să fie "line", ceea ce tradus în limbaj matematic înseamnă că liniile poligonale nu sînt decît niste aproximări empirice ale unei curbe, care reprezintă "legea" de distribuție a populației în raport cu variabila analizată.

În statistica matematică există un arsenal de funcții care descriu asemenea distribuții ideale ale colectivităților statistice, după mai multe tipuri de curbe. Dintre acestea, cea mai cunoscută - și probabil cu aplicabilitatea cea mai largă - este așa-numita curbă a lui Gauss, sub formă de clopot, cunoscută și sub numele de curbă sau lege normală de distribuție. Două probleme importante se pun în legătură cu trecerea de la un set de date individuale (valori pentru indivizi empirici) la asemenea curbe matematice:

Mai întîi, una practică, și anume, avînd o distribuție empirică, se urmărește să se vadă dacă se poate afirma că această distribuție poate fi considerată o aproximație suficient de bună a uneia teoretice, mai exact trebuie testat dacă frecvențele empirice pot fi privite cu particularizări ale celor teoretice. Se consideră că lotul de indivizi studiați reprezintă un eșantion reprezentativ al unei populații, potențial infinite, iar testarea constă în calcularea semnificației diferențelor dintre frecvențele empirice și cele care ar rezulta, la același volum de populație și la aceleași intervale de valori, dacă distribuția urmează o curbă matematică ideală. Este deci o chestiune de testare a semnificației diferențelor, legată de eșantionare, despre care va fi vorba în ultima parte a cursului.

A doua problemă este una teoretică, de principiu, cerînd să se indice calea inversă, cea prin care s-ar putea trece de la valorile indivizilor empirici la o curbă. Presupunînd că popula-



ția este foarte mare și că valorile sînt măsurate cu o precizie înaltă, spre a putea discrimina bine indivizii, procedura care s-ar impune, la prima vedere, este următoarea. Se vor înjumătăți intervalele de grupare (dublîndu-se deci numărul lor) și se va construi un nou poligon al frecvențelor. Este evident că de această dată linia poligonală obținută nu va mai prezenta variații așa de mari în orientarea segmentelor vecine. (Acestea se vor îmbina prin unghiuri mai puțin ascuțite). S-ar părea că dacă se continuă indefinit acest procedeu - evident, în principiu - trecerile vor deveni tot mai line, urmînd ca la limită să se ajungă la o curbă. Numai că această procedură nu este posibilă nici măcar pe planul intuitiv. Intr-adevăr, prin înjumătățirea continuă a intervalor, frecvența fiecărei clase descrește (în medie, cu fiecare pas frecvențele se reduc la jumătate) iar la limită tind spre zero, poligonul frecvențelor căzînd pur și simplu peste axa absciselor. Pentru a scăpa din această capcană a infinitului mare și mic, trebuie mers pe o cale diferită, deschisă de utilizarea frecvențelor sub o altă formă decît cea prezentată în tabelele anterioare: este vorba de așa-numitele frecvențe cumulate. Uneori această procedură este folosită și pentru scopuri empirice, așa că ne vom opri puțin asupra ei.

S-a văzut că, în cazul unei variabile statistice, pentru fiecare valoare  $x_i$  s-a pus în corespondență o frecvență  $k_i$ , reprezentînd mulțimea indivizilor care primesc valoarea respectivă, după caracteristica studiată. Ne putem întreba uneori și cîți indivizi dintr-o populație iau valori sub sau peste o anumită valoare dată. Pentru a putea răspunde la întrebarea de mai sus, indiferent despre ce valoare ar fi vorba, tabelul repartiției de frecvență trebuie modificat, cumulînd toate frecvențele pînă la fiecare  $x_i$  sau peste fiecare  $x_i$ . Se obține astfel două tipuri de tabele de frecvențe cumulate. De exemplu, pentru datele din tabelul 7, cumulara frecvențelor va conduce la situația din tabelul 9.



Mediile ( $x_i$ )	Frecvențe cumulate			
	A. Pînă la $x_i$		B. Peste $x_i$	
	Absolute	%	Absolute	%
2	0	0	600	100
3	30	5,0	570	95,0
4	70	11,7	530	88,3
5	200	33,3	400	66,7
6	350	58,3	250	41,7
7	450	75,0	150	25,0
8	525	87,5	75	12,5
9	575	95,8	25	4,7
10	600	100	0	0

**Tabelul 9.** Frecvențele cumulate ale caracteristicii prezentate în tabelul 7

Se observă imediat utilitatea tabelelor de acest gen. Ele oferă posibilitatea de a citi rapid cîți indivizi sau ce procent dintre ei se află sub un anume prag sau îl depășesc. De exemplu, în cazul de față 200 din cei 600 de candidați (33,3%) n-au reușit să obțină media 5; sau 75 din 600 (12,5%) au reușit să depășească media 8.

Observație. În tabelul 9 am trecut pe coloana întâi doar valorile, deși caracteristica era constituită prin intervale de grupare. Se observă, din felul cum au fost scrise aceste intervale, că de fapt porțiunea B a tabelului conține frecvențe cumulate reprezentînd numărul de indivizi care iau nu numai valori mai mari decît  $x_i$  dar și pe  $x_i$ , inclusiv. Tot așa de bine valorile întregi 3,4,5... puteau constitui capetele superioare ale intervalelor și atunci situația se inversa.

Prin reprezentarea grafică a frecvențelor din tabelul 9, cu ajutorul poligonului frecvențelor, se obține figura 7.



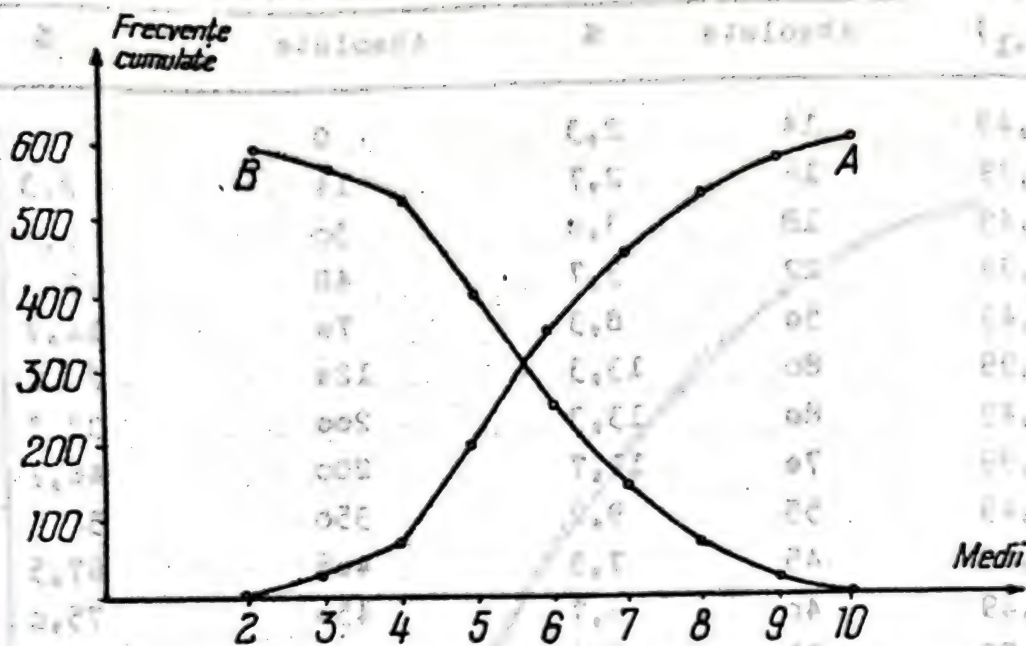


Fig. 7 Poligonul frecvențelor cumulate din tabelul 9

Este evident că întotdeauna frecvențele cumulate de tip A vor conduce la o linie poligonală ascendentă, în timp ce cele de tip B la una decendentă. Inclinația fiecărui segment față de axa orizontală este cu atât mai mare cu cât numărul indivizilor care se adaugă (sau se scad) de la o valoare la următoarea este mai mare. În cazul de față, unghiul cel mai mare cu axa Ox îl fac segmentele ce unesc punctele corespunzătoare valorilor 5 și 6, iar cel mai mic cele de la capetele liniei.



Mediile ( $x_i, x_{i+1}$ )	Frecvențe obișnuite		Frecvențe cumulate, tip A	
	Absolute	%	Absolute	%
2 - 2,49	14	2,3	0	0
2,50 - 2,99	16	2,7	14	2,3
3,00 - 3,49	18	3,0	30	5,0
3,50 - 3,99	22	3,7	48	8,0
4,00 - 4,49	50	8,3	70	11,7
4,50 - 4,99	80	13,3	120	20,0
5,00 - 5,49	80	13,3	200	33,3
5,50 - 5,99	70	11,7	280	46,7
6,00 - 6,49	55	9,2	350	58,3
6,50 - 6,99	45	7,5	405	67,5
7,00 - 7,49	40	6,7	450	75,0
7,50 - 7,99	35	5,8	490	81,7
8,00 - 8,49	30	5,0	525	87,5
8,50 - 8,99	20	3,3	555	92,5
9,00 - 9,49	15	2,5	575	95,8
9,50 - 10	10	1,7	590	98,3
TOTAL	600	100	600	100

Tabelul 10. Frecvențele simple și cumulate, obținute prin înjumătățirea intervalului de grupare (cazul prezentat în tabelul 7)

Revenind acum la problema curbelor de frecvențe, se poate constata ușor că procedura de înjumătățire a intervalelor nu ne mai conduce la impas dacă folosim frecvențele cumulate. Să luăm în considerare numai tipul A și să reluăm exemplul anterior, presupunând că prin înjumătățirea intervalelor se obține datele din tabelul 10, tabel în care s-au trecut și frecvențele cumulate de tip A (până la  $x_i$ , adică sub valoarea din stînga a intervalului).

În figura 8 am reluat, pentru comparație, linia poligonală A din figura 7, căreia i-am alăturat o linie poligonală construită cu frecvențele cumulate din tabelul 10. Se observă imediat cum unghiurile dintre segmentele vecine se "îndulcesc" și cum, la li-



mită, linia poligonală tinde spre o curbă de genul celei trasate pe aceeași figură.

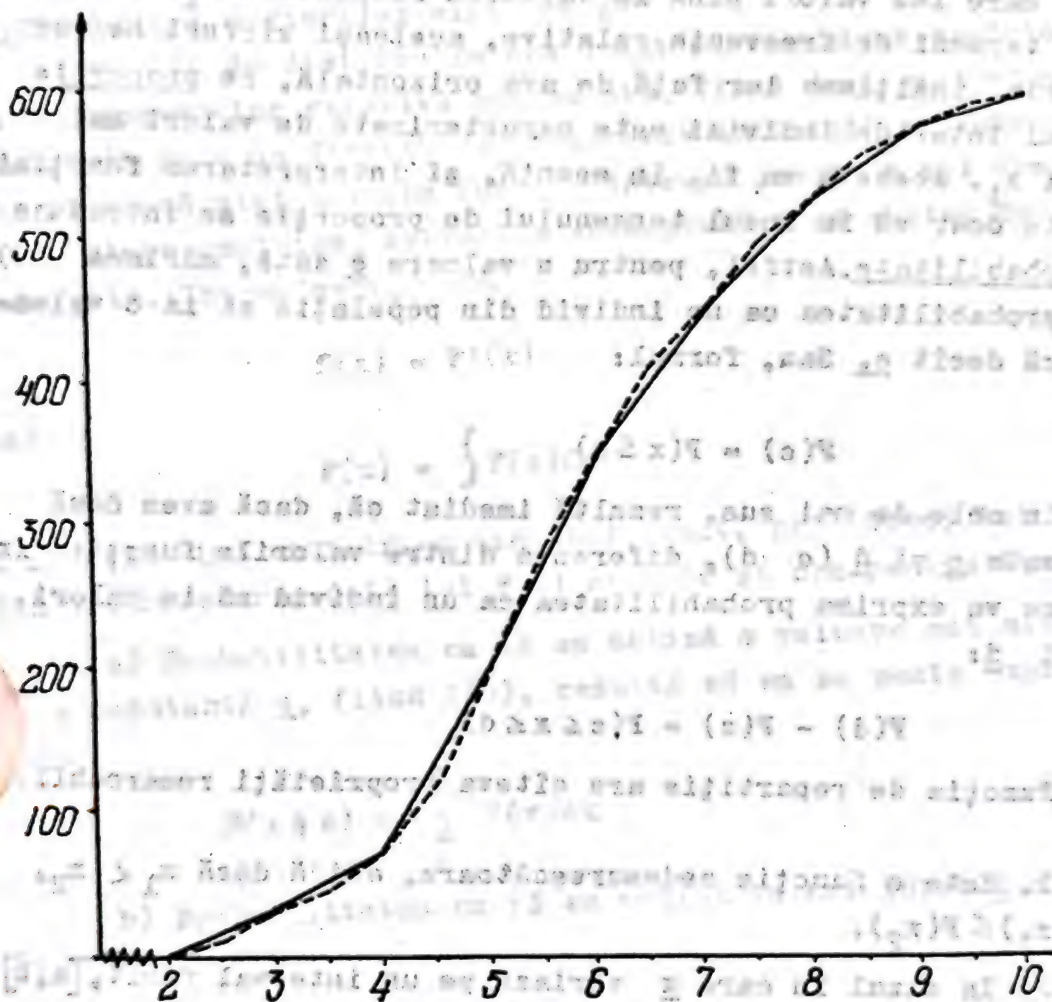


Fig. 8 Frecvențele cumulate de tip A din Fig. 7 (linia continuă) și cele obținute prin înjumătățirea intervalelor (linia punctată)

Operația descrisă mai sus se poate realiza atât cu frecvențe absolute cât și cu procente sau frecvențe relative (la unitate). Să presupunem că am utilizat aceste din urmă mărimi și că am reușit să facem saltul de la linii poligonale la o curbă. Am trecut atunci din domeniul statisticii descriptive în cel al matematicii, mai exact în cel al teoriei probabilităților. Curbă obținută va fi imaginea grafică a unei funcții  $F(x)$ , care poartă denu-



mirea de funcție de repartiție a unei variabile aleatoare  $X$ . Să vedem care este interpretarea acestei funcții. Așa cum am arătat, vîrfurile poligonului frecvențelor cumulate ne indică volumul indivizilor care iau valori pînă la valoarea respectivă  $x_i$ . Dacă raționăm în termeni de frecvențe relative, aceleași vîrfuri ne vor indica, prin înălțimea lor față de axa orizontală, ce proporție din numărul total de indivizi este caracterizată de valori mai mici decît  $x_i$ . Aceasta va fi, în esență, și interpretarea funcției  $F(x)$ , atîta doar că în locul termenului de proporție se introduce cel de probabilitate. Astfel, pentru o valoare  $c$  dată, mărimea  $F(c)$  ne arată probabilitatea ca un individ din populație să ia o valoare mai mică decît  $c$ . Sau, formal:

$$F(c) = P(x \leq c)$$

Din cele de mai sus, rezultă imediat că, dacă avem două valori fixate  $c$  și  $d$  ( $c < d$ ), diferența dintre valorile funcției în  $d$  și  $c$  ne va exprima probabilitatea ca un individ să ia valori între  $c$  și  $d$ :

$$F(d) - F(c) = P(c \leq x \leq d)$$

Funcția de repartiție are cîteva proprietăți remarcabile evidente:

1. Este o funcție nedescrescătoare, adică dacă  $x_1 < x_2$ , atunci  $F(x_1) \leq F(x_2)$ .

2. În cazul în care  $x$  variază pe un interval finit,  $[a, b]$ ,

$$F(a) = 0 \text{ și } F(b) = 1$$

ceea ce corespunde faptului simplu că nici un individ nu ia valori mai mici decît  $a$  și mai mari decît  $b$ .

3. Dacă intervalul de variație este  $(-\infty, +\infty)$  atunci curba va tinde asimptotic la 0, spre  $-\infty$ , și la 1, spre  $+\infty$ .

Vorbind despre linia poligonală a frecvențelor cumulate, am observat că înclinarea acestora în raport cu axa  $Ox$  este în funcție de proporția indivizilor ce se adaugă de la o valoare la următoarea, respectiv de proporția indivizilor cuprinși în intervalul de grupare în cauză. Acest lucru se reflectă și în cazul funcției  $F(x)$ ; dacă luăm două puncte de curbă și le unim printr-o



coardă, la limită, coardele respective devin tangente la curbă, tangente care, așa cum se știe, indică "viteza de creștere" a funcției  $F(x)$  într-un punct. Se ajunge astfel imediat la o altă funcție  $f(x)$ , derivata lui  $F(x)$ , care poartă numele de funcție de distribuție a variabilei aleatoare  $X$ .

Funcția de distribuție are drept corespondent empiric poligonul frecvențelor relative necumulate, fiind deci o ilustrare a modului în care se distribuie cantitativ indivizii pe intervalul  $(a, b)$ . Numai că aici valoare funcție într-un punct nu mai are nici o interpretare empirică. Astfel de interpretări se obține doar pe subintervale. Într-adevăr, dacă

$$f(x) = F'(x)$$

atunci

$$F(x) = \int f(x) dx$$

deci funcția  $F(x)$  este integrala (mai exact, primitiva) lui  $f(x)$  și din interpretarea dată lui  $F(x)$  rezultă și cea pentru  $f(x)$ :

a) Probabilitatea ca să se obțină o valoare mai mică decât o constantă  $c$ , fiind  $F(c)$ , rezultă că ea se poate exprima și astfel:

$$P(x \leq c) = \int_a^c f(x) dx$$

b) Probabilitatea ca să se obțină valori între  $c$  și  $d$

va fi:

$$P(c \leq x \leq d) = \int_c^d f(x) dx$$

c) Cum  $F(b) = 1$ , rezultă că integrala lui  $f(x)$  pe tot intervalul  $(a, b)$  este egală cu unitatea:

$$\int_a^b f(x) dx = 1$$

relația fiind valabilă și când limitele intervalului sînt finite.



După cum se știe, o interpretare grafică a integralei unei funcții o constituie aria figurii delimitate de curbă, axa  $Ox$  și două drepte verticale duse prin  $c$  și  $d$  - limitele integralei. O asemenea situație este marcată în figura 9. Aria punctată reprezintă probabilitatea ca un individ să ia valori mai mici decât  $c$ , iar cea hașurată ca să ia valori între  $c$  și  $d$ . Sau, vorbind în termeni empirici, aria punctată este egală cu proporția indivizilor care nu depășesc valoarea  $c$ , pe când cea hașurată proporția ponderea celor cu valori între  $c$  și  $d$ .

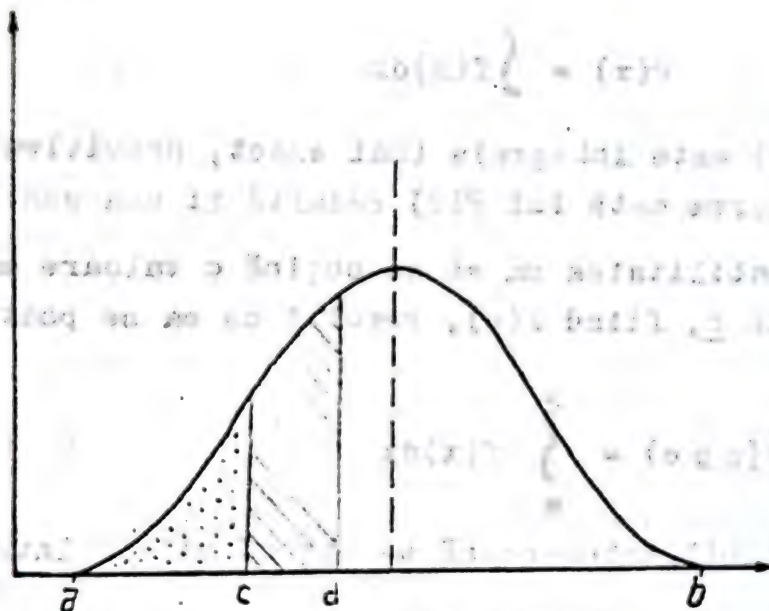


Fig. 9 O curbă simetrică

Diversele distribuții empirice conduc <sup>spre</sup> câteva modele teoretice de curbe, dintre care mai frecvent întâlnite sînt:

1. Curbele simetrice, ca cea din fig.9., în care maximum frecvențelor se atinge la mijlocul intervalului  $(a,b)$  sau, dacă intervalul este  $(-\infty, +\infty)$ , frecvențele ating o zonă de maxim. scăzînd simetric spre stînga și spre dreapta.

2. Curbele moderat asimetrice. Zona de frecvențe maxime se află deplăsată spre un capăt al intervalului și apoi frecvențele descresc cu viteze inegale într-o parte și cealaltă.



3. Curbele în formă de J. Reflectă situațiile în care maximum de cazuri se află într-un capăt al intervalului, crescînd constant spre celălalt.

4. Curbele în formă de U. Descriu cazurile de distribuții în care indivizii se concentrează spre cele două capete ale intervalului, avînd frecvențe minime la mijlocul acestuia.

Dintre curbele simetrice, cea mai cunoscută este curba lui Gauss, definită pe intervalul  $(-\infty, +\infty)$ . Experiențele empirice care pot duce la distribuții normale sînt foarte multe. Amintesc doar două clasice. Să presupunem că cineva aruncă o bilă în intenția de a o plasa pe o linie trasată la o oarecare distanță. Măsurînd de fiecare dată distința de la punctul de impact al bilei cu solul la linia respectivă și atribuind semnul "+" valorilor ce corespund unor aruncări lungi și "-" celor scurte, va obține o distribuție de frecvențe (indivizii fiind aruncările, iar valorile aceste distanțe cu semn) pe intervale de grupare, distribuție care se va dovedi simetrică în jurul lui 0. O distribuție asemănătoare se obține și în cazul unor măsurări repetate a lungimii unei bare cu un același instrument, suficient de fin pentru a putea evidenția diferențe foarte mici de măsurare. În ambele exemple amintite valorile variabilelor pot fi interpretate ca erori, de aruncare, respectiv, de măsurare. În general, se consideră că erorile care se datorează unor factori întîmplători, nesistematici, urmează o distribuție normală în jurul lui zero, respectiv măsurătorile afectate doar de astfel de erori se distribuie normal în jurul valorii "adevărate" a obiectului de măsurat. De altfel, curba normală mai poartă denumirea și de curbă a erorilor.

Desigur că nu toate curbele normale sînt identice; ele diferă în esență prin două elemente: valoarea în care se găsește punctul central și gradul de "împrăștiere" a indivizilor statistici în jurul acestei valori. De pildă, în primul caz amintit valoarea centrală va fi nulă iar, în al doilea, egală cu dimensiunea fizică de măsurat. De asemenea, în prima situație abaterile de la punctul de simetrie vor fi incomparabil mai mari decît în al doilea.

Datorită cîmpului foarte larg de aplicabilitate, virtuțile curbei normale sînt uneori prezentate exagerat, în sensul că



se consideră despre multe fenomene ca ar urma o asemenea distribuție, fără a avea suficiente temeiuri pentru atare supoziție. Astfel, în domeniul științelor sociale, se presupune adeseori, tacit sau explicit, că o serie de însușiri ale populațiilor umane sînt normal distribuite. Numai din faptul că valorile extreme sînt rare nu trebuie să tragem concluzia că distribuția este simetrică și, cu atît mai puțin, normală.

Voi încheia aceste sumare considerații asupra curbelor statistice cu precizarea că interesul față de ele nu constă numai sau în principal în funcția lor ilustrativă. Dacă se constată că o distribuție empirică poate fi suficient de bine apropiată de una teoretică atunci se pot extrapola asupra primeia toate cunoștințele pe care le avem despre a doua. Dacă, în plus, prin cercetări repetate se constată mereu o asemenea apropiere, curba teoretică va reflecta ceea ce este constant în fenomene, o legitate sau regularitate socială, aflată în spatele unei mase de evenimente întîmplătoare.



## Cap. VIII. INDICATORI AI CARACTERISTICILOR CANTITATIVE

Informația surprinsă sub forma unei variabile statistice, mai ales cînd numărul valorilor este mare, poate fi cu greu sesizată de cercetător sau de beneficiarul rezultatelor sale și astfel se naște nevoia concentrării ei sub forma unor mărimi unice, care să reflecte, sintetic, întreaga situație descrisă de o caracteristică. Desigur, o atare concentrare nu are numai rolul pragmatic sugerat aici, ci de cele mai multe ori, prin ea se realizează și un scop gnoseologic major: mărimile unice calculate pentru caracteristica în cauză sînt chemate să reflecte ceea ce este general, tipic, normal, constant etc. în fenomenul social analizat. Mai mult, calculul acestor mărimi va constitui o condiție necesară pentru continuarea analizelor la situații mai complexe, mai exact, pentru trecerea de la forma descriptiv-statistică la cea care ne ajută să explicăm fenomenele obiective descrise statistic.

În funcție de genul de informație pe care o reprezintă, de ceea ce urmăresc să surprindă în cadrul unei distribuții, mărimile respective, numite generic indicatori ai caracteristicilor, sînt de regulă, clasificate în mai multe categorii. Dintre acestea, în manualul de față, o să prezint doar două: categoria indicatorilor de poziție și cea a indicatorilor de împrăștiere.

### 1. Indicatori de poziție

Clasa acestor indicatori cuprinde mărimile ce se calculează pentru caracteristicile cantitative, cu scopul de a surprinde, în mod foarte concentrat, nivelul valoric mijlociu al grupului de indivizi. Altfel spus, date fiind valorile luate de populația noastră, de la  $x_1$  (cea mai mică) la  $x_g$  (cea mai mare), problema care se pune este de a determina o valoare care să caracterizeze cel mai bine ansamblul de indivizi, o valoare în jurul căreia se grupează celelalte, o valoare care să poată constitui un fel de etalon, de caz tipic pentru seria statistică a variabilei  $X$ . Atenționez că o astfel de mărime poate coincide efectiv cu una dintre valorile



$x_1, \dots, x_s$  luate de indivizii noștri sau nu va coincide. În orice caz, ea trebuie să se găsească undeva în interiorul intervalului  $(x_1, x_s)$  și, de regulă, într-o poziție centrală a acestuia.

Există evident mai multe posibilități de a construi un asemenea indicator, lucru vizibil și din cerințele pe care el este chemat să le îndeplinească, cerințe formulate, recunosc, destul de vag în rândurile precedente. Dar, cu tot caracterul vag al exprimării se poate vedea că nu toate cerințele sînt echivalente. Astfel, de pildă, valoarea mijlocie poate să nu fie tipică, în sensul că ea nu caracterizează nici un individ din populație sau că în jurul ei nu găsim o concentrare deosebită a unităților noastre statistice. Astfel de situații se întîlnesc atunci cînd indivizii se plasează cu frecvențe mai mari spre capetele intervalului de variație a valorilor (se distribuie după o curbă în formă de U), valoarea mijlocie fiind mai slab populată.

Dintre indicatorii de poziție analizați în literatura statistică, voi prezenta doar trei, cei mai cunoscuți și mai des utilizați: media (aritmetică), mediana și modul (valoarea modală) în ordinea în care i-am enumerat.

### 1.1. Media

După cum se știe, există mai multe tipuri de mărimi care poartă denumirea de medie, dintre care cea mai răspîdită în analizele statistice este așa-numita medie aritmetică. Dat fiind faptul că nu voi aborda și celelalte genuri (media geometrică, media armonică), în cele ce urmează voi utiliza simplu denumirea de medie sau valoare medie, înțelegîndu-se că este vorba de cea aritmetică.

Noțiunea este foarte cunoscută pentru că este adesea folosită în viața de zi cu zi. Un student, care într-o sesiune ia, în ordinea examenelor date, următoarele note: 9, 10, 8, 8, 10, va spune imediat că a terminat sesiunea cu media 9, înlocuind prin această informație sintetică informația mai detaliată care s-ar fi transmis prin comunicarea celor 5 note; de asemenea, comunicîndu-se media 9, avem, implicit și direct, o imagine generală asupra nivelului mijlociu la care a fost apreciată pregătirea sa.



Calculul mediei, în cazul simplu ca cel de mai sus, nu ridică, evident, nici o problemă; se adună notele (deci valorile caracteristicii) și se împarte la numărul cazurilor de intervenție (numărul indivizilor), chit că în situația dată studentul nostru "vede" imediat media, folosind una din proprietățile acesteia.

În general, procedura formală de definire a indicatorului este următoarea. Având dată o mulțime de  $n$  indivizi, fiecare dintre aceștia luând în raport cu o variabilă  $X$  o valoare  $x_i$  ( $i=1,2,\dots,n$ ) atunci se definește media variabilei  $X$ , notată cu  $\bar{x}$ , prin relația:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

sau, prescurtat:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Dacă populația statistică este mare, o aceeași valoare va fi luată de mai mulți indivizi și caracteristica noastră apare sub forma cunoscută a unei distribuții de frecvență. Atunci formula mediei se va scrie:

$$\bar{x} = \frac{k_1 x_1 + k_2 x_2 + \dots + k_g x_g}{n}$$

sau, prescurtat:

$$\bar{x} = \frac{\sum_{i=1}^g k_i x_i}{n} \quad (2)$$

unde notațiile sînt cele din capitolul precedent:  $x_1, x_2, \dots, x_g$  sînt cele  $g$  valori distincte luate efectiv (sau, dacă variabila s-a construit prin intervale de grupare, cele  $g$  valori mijlocii ale fiecărui interval) iar  $k_1, k_2, \dots, k_g$  - frecvențele cu care sînt luate valorile respective.



Formula (1) poartă, în general, denumirea de medie aritmetică simplă, pe cînd (2), de medie ponderată. În fapt, este vorba de același lucru, deci nu sînt două medii de specii diferite, căci în (2) se concentrează în produse sume parțiale din (1). Intr-adevăr, dacă, de exemplu, 10 indivizi statistici iau fiecare valoarea 1500, atunci suma valorilor pentru ei se poate obține adunînd pe 1500 cu el însuși de 10 ori sau mai simplu înmulțindu-l cu 10. Revenind la cazul studentului nostru, media notelor sale se poate calcula fie:

$$\bar{x} = \frac{9+10+8+8+10}{5} = 9$$

fie

$$\bar{x} = \frac{2.8+1.9 + 2.10}{5} = 9$$

În domeniul sociologic, cele mai frecvente situații sînt cele care fac apel la formula (2), căci se lucrează cu populații mari. Pentru acest motiv, voi folosi mai frecvent în continuare formula (2), deși, repet, între (1) și (2) nu este decît o diferență de scriere: în primul caz, indicele 1, după care se fac însumarea, urmărește indivizii, adică valorile individuale scrise într-o anumită ordine și în număr de n, pe cînd, în al doilea, valorile distincte după care se constituie cele s clase ale caracteristicii, așezate, în general, de la cea mai mică la cea mai mare.

Pentru ilustrarea modului practic de calculare a mediei, să reluăm, mai întîi, caracteristica prezentată în capitolul precedent, în tabelul 6 (distribuția gospodăriilor după numărul membrilor ce le compun) - În scopul facilității calculelor, datele scrise în tabele pe orizontală se trec acum pe verticală; se transpun cele două linii în coloane, una conținînd valorile, alta frecvențele. Se mai adaugă apoi o a treia coloană ce rezultă înmulțind fiecare frecvență cu valoarea corespunzătoare, ceea ce înseamnă că suma de la numărătorul formulei (2) se obține adunînd cifrele de pe această din urmă coloană. Pentru cazul nostru, se va proceda astfel:



Valori $x_1$	Frecvențe $k_1$	Produse $k_1 x_1$
1	30	30
2	40	80
3	60	180
4	50	200
5	30	150
6	20	120
7	10	70
8	5	40
9	5	45
Total ( $\Sigma$ )		915

Deci:

$$\bar{x} = \frac{915}{250} = 3,66$$

Adică mărimea medie a unei gospodării din localitatea respectivă este de 3,66 persoane. Ne aflăm aici într-un caz în care media nu coincide cu nici una din valori și nici n-ar putea coincide decât dacă printr-o șansă deosebită, ar ieși sub formă de număr natural. Chiar dacă este absurd să vorbim de o gospodărie cu 3,66 persoane, aceasta nu ne împiedică deloc să folosim mărimea medie pentru a indica nivelul mijlociu al dimensiunii gospodăriilor din satul studiat.

Observație. Suma  $\Sigma k_1 x_1$ , adică suma valorilor tuturor unităților din populația studiată, poate uneori avea interpretări foarte directe și ea poate fi cunoscută fără a recurge la calculele de mai sus. De pildă, în exemplul de față, 915 nu este decât numărul total de indivizi ai satului, ce putea fi obținut fără a construi propriu-zis o caracteristică; deci și media poate fi uneori aflată fără detalierea datelor. Dacă, un alt caz, știm că într-o întreprindere fondul de retribuție pe o lună este de 1.500.000 lei și întreprinderea are 500 de angajați, este evident că retribuția medie pe lucrător va fi de 3.000 lei, fără a ști exact



Valori	Frecvențe	Produse
$x_1$	$k_1$	$k_1 x_1$

1	30	30
2	40	80
3	60	180
4	50	200
5	30	150
6	20	120
7	10	70
8	5	40
9	5	45

Total (C) 250 915

Deci:

$$\bar{x} = \frac{915}{250} = 3,66$$

Adică mărimea medie a unei gospodării din localitatea respectivă este de 3,66 persoane. Ne aflăm aici într-un caz în care media nu coincide cu nici una din valori și nici n-ar putea coincide decât dacă printr-o șansă deosebită, ar ieși sub formă de număr natural. Chiar dacă este absurd să vorbim de o gospodărie cu 3,66 persoane, aceasta nu ne împiedică deloc să folosim mărimea medie pentru a indica nivelul mijlociu al dimensiunii gospodăriilor din satul studiat.

Observație. Suma  $\sum k_1 x_1$ , adică suma valorilor tuturor unităților din populația studiată, poate uneori avea interpretări foarte directe și ea poate fi cunoscută fără a recurge la calculele de mai sus. De pildă, în exemplul de față, 915 nu este decât numărul total de indivizi ai satului, ce putea fi obținut fără a construi propriu-zis o caracteristică; deci și media poate fi uneori aflată fără detalierea datelor. Dacă, un alt caz, știm că într-o întreprindere fondul de retribuție pe o lună este de 1.500.000 lei și întreprinderea are 500 de angajați, este evident că retribuția medie pe lucrător va fi de 3.000 lei, fără a ști exact



cum se distribuie retribuițiile efective.

În situațiile în care produsele  $k_i x_i$  duc la valori mari sau presupun calcule mai greoaie din pricina zecimalelor, se poate aplica o formulă de calcul simplificat a mediei. Se pleacă de la constatarea că dacă din fiecare valoare a caracteristicii se scade o mărime constantă și se calculează media cu diferențele rămase, pentru a se reveni la media adevărată <sup>trebuie</sup> să i se adauge prima valoare constantă scăzută. Sau, scris riguros:

$$\bar{x} = \frac{\sum_{i=1}^s k_i (x_i - a)}{n} + a \quad (3)$$

unde  $a$  este constanta menționată. Se poate arăta ușor că formulele (2) și (3) sînt echivalente. Să aplicăm formula (3) pentru situația prezentată în tabelul 7 din capitolul anterior. Era vorba acolo de distribuția numărului de candidați la admitere pe intervale de note: 2 - 2,99 , 3 - 3,99 , ..., 9 - 10. Să atașăm fiecărui interval drept valoare caracteristică notele 2,50, 3,50, ..., 9,50, care devin astfel valorile variabilei statistice. Pentru a evita înmulțirea cu zecimale, vom alege ca valoare ajutătoare:

$$a = 5,50$$

Atunci, calculul se ordonează astfel:

$x_i$	$k_i$	$x_i - a$	$k_i(x_i - a)$
2,50	30	-3	90
3,50	40	-2	80
4,50	130	-1	130
5,50	150	0	0
6,50	100	1	100
7,50	75	2	150
8,50	50	3	150
9,50	25	4	100
Total			600

$$500 - 300 = 200$$



După cum se vede, în ultima rubrică am aranjat cifrele pe două coloane: cele pozitive în stînga și cele negative în dreapta, pentru a facilita însumarea algebrică. S-a obținut că suma de la numărătorul fracției este 200, valoare ce împărțită la cei 600 de indivizi ne va da 0,33. Deci

$$\bar{x} = 0,33 + 5,50 = 5,83$$

Aceasta este deci nota medie obținută de candidații la admitere, în cazul ipotetic prezentat.

Observație. În cazul caracteristicilor cu intervale de grupare, media care se obține este doar o aproximație a mediei adevărate, pentru calculul căreia ar fi trebuit luată în considerare valoarea exactă a fiecărui individ. De regulă, aproximația astfel obținută este satisfăcătoare, deoarece erorile ce se introduc prin alegerea valorilor caracteristice pentru intervale se compensează de-a lungul scalei intervalelor. Astfel, dacă distribuția este simetrică sau moderat asimetrică, pe intervalele din stînga indivizii au tendința de a se grupa mai mult spre capătul din dreapta (deci mijlocul intervalului subestimează media valorilor din interval), pe cînd în partea superioară a scalei, datorită descreșterii frecvențelor, valorile caracteristice alese supraestimează mediile parțiale pe intervale.

#### Proprietăți ale mediei

1. În calitatea sa de mărime mijlocie, este evident că media trebuie să se găsească plasată pe scala valorilor între cea mai mică și cea mai mare valoare:

$$x_1 < \bar{x} < x_n$$

Proprietatea poate primi ușor o demonstrație riguros matematică.

2. Media este o mărime exprimată în aceleași unități de măsură ca și variabila pe care o reprezintă. Astfel, dacă valorile variabilei sînt note, ani, unități monetare etc., media va ilustra același gen de mărimi.

3. Calitatea de mărime mijlocie a mediei poate fi exprimată și prin proprietatea că suma abaterilor valorilor individuale



de la medie este nulă. Sau, altfel spus, abaterile valorilor din stînga mediei se compensează cu ale celor din dreapta ei. Formal proprietatea poate fi astfel exprimată:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

pentru cazul în care se aplică formula (1) sau:

$$\sum_{i=1}^s k_i (x_i - \bar{x}) = 0$$

pentru formula (2).

Intr-adevăr, dacă luăm, de pildă, relația a doua și desfacem parantezele, egalitatea devine echivalentă cu:

$$\sum_{i=1}^s k_i x_i - \sum_{i=1}^s \bar{x} \cdot k_i = 0$$

Sau, dînd factor comun pe  $\bar{x}$  în termenul al doilea:

$$\sum_{i=1}^s k_i x_i - \bar{x} \sum_{i=1}^s k_i = 0$$

Împărțind fiecare termen al relației cu  $n$ , rezultă:

$$\frac{\sum_{i=1}^s k_i x_i}{n} - \frac{n \cdot \bar{x}}{n} = 0$$

Primul termen al diferenței este exact valoarea medie dată de (2), iar în termenul care se scade, după simplificarea cu  $n$ , rămîne același lucru:

$$\bar{x} - \bar{x} = 0$$



și formula este demonstrată.

4. Tot ca o expresie a calității de mărime mijlocie, media posedă și proprietatea că, înmulțită cu numărul indivizilor, se obține exact suma valorilor individuale. Faptul rezultă imediat dacă formulele (1) și (2) se înmulțesc cu  $n$ :

$$\sum_{i=1}^n x_i = n \cdot \bar{x}$$

$$\sum_{i=1}^s k_i x_i = n \cdot \bar{x}$$

Proprietatea reflectă faptul că media este acea valoare pe care ar lua-o indivizii dacă toți ar fi identici din punctul de vedere al caracteristicii respective. De pildă, suma notelor studentului despre care am pomenit la început este de 45, sumă la care s-ar fi ajuns dacă studentul în cauză ar fi luat la toate cele cinci examene nota 9. Sau, în alt exemplu folosit, dacă toți cei 500 de lucrători ar fi avut aceeași retribuție - 3.000 lei - cât reprezintă media pe întreprindere, fondul total de retribuție ar fi același.

5. Media aritmetică rămâne neschimbată dacă frecvențele se înmulțesc sau se împart cu un același număr. Proprietatea este evidentă, căci amplificând fiecare frecvență  $k_i$  cu o valoare  $m$ , aceasta din urmă poate fi dată factor comun la numărător; la fel și numărul indivizilor de la numitor (care nu este decât suma frecvențelor) se va înmulți cu  $m$  și, prin urmare, fracția se va simplifica imediat cu  $m$ . Această proprietate ar putea fi utilă în calcule în situațiile când frecvențele au un multiplu comun ce poate fi dat factor și deci cu care se va simplifica fracția. Din păcate, în calculele pe date reale, astfel de situații sînt improprie de întîlnit.

În contextul acestei proprietăți, mai menționez că media aritmetică poate fi scrisă și cu ajutorul frecvențelor relative. Împărțind fiecare termen de la numărătorul fracției din formula (2) cu  $n$  și deci eliminînd numitorul, se ajunge la:



$$\bar{x} = \sum_{i=1}^n p_i x_i \quad (2')$$

6. Dacă înmulțim sau împărțim valorile caracteristicii cu același număr, valoarea medie se înmulțește sau se împarte și ea tot cu numărul respectiv. Proprietatea rezultă și ea imediat tot din formula de definiție a mediei. Coroborînd această proprietate cu cea privind translația valorilor, folosită în formula (3) de simplificare a calculelor, ajungem la concluzia că dacă valorile variabilei X suferă o transformare lineară, atunci media noii variabile, X', se regăsește prin aceeași transformare. Adică, dacă

$$X' = a + b \cdot X$$

atunci

$$\bar{x}' = a + b \cdot \bar{x}$$

După cum se știe, media este o mărime foarte curent folosită în diferite domenii ale cunoașterii. Uneori utilizarea sa este abuzivă, mai cu seamă atunci cînd ea este luată drept valoare mijlocie sau chiar tipică și, de fapt, nu se întîmplă acest lucru. Intr-adevăr, din formula (2') se vede foarte bine că media este în funcție de doi factori deosebiți: valorile caracteristicii și ponderile pe care le au indivizii ce iau aceeași valoare în populație. Se poate întîmpla ca asupra mărimei medii să aibă o influență sensibilă una sau cîteva valori foarte mari ori foarte mici, dar care caracterizează un număr restrîns de unități ale populației. Problema se va înțelege mai bine dacă dăm un exemplu foarte simplu. Să presupunem că într-o echipă de muncă lucrează patru muncitori necalificați care, în funcție de vechimea lor primesc o retribuție lunară de 2500, 2600, 2700 și respectiv 2800 lei. Alături de ei se mai află și un maestru cu o retribuție lunară de 4400 lei. Făcînd media retribuției pe întreaga echipă se va obține:

$$\bar{x} = \frac{2500+2600+2700+2800+4400}{5} = \frac{15.000}{5} = 3000$$



Deci, dacă, în medie, un membru al echipei respective câștigă 3000 lei lunar, în realitate patru din cinci se află sub această valoare și doar unul singur (mult) peste ea. În asemenea cazuri, media nu redă evident o valoare mijlocie a grupului, în sensul intuitiv al termenului de mijlociu, reflectat mai bine de un alt indicator: mediana.

Comentînd situații asemănătoare cu cea de mai sus, autorii unui manual de statistică dedicat economiștilor (Vezi M.Biji și E.Biji, 1979, pg.101) merg pînă acolo încît recomandă calcularea mediei doar pentru situații omogene, tocmai pentru a se evita interpretări greșite. Cred că se exagerează aici puțin; din cauza unor utilizări și interpretări greșite nu se poate ajunge la a interzice calcularea unui indicator.

Important pentru cel care folosește indicatorii statistici, media în cazul de față, este să înțeleagă semnificația exactă a valorii calculate și să o utilizeze exclusiv în scopurile care se potrivesc acestei semnificații. Mediile nu au un caracter fictiv, nu sînt lipsite de conținut, nu sînt neprezentative, în situații ca cea de mai sus, cum lasă să se înțeleagă autorii menționați. Ele pur și simplu nu ne dau ceea ce unii așteaptă de la ele sau pretind că ne dau. Cerința omogenității are desigur o anumită rațiune. Media, am spus, nu poate să reflecte decît o parte din informația surprinsă în caracteristică și este evident că, cu cît populația este mai omogenă, cu atît media va "trăda" mai puțin din cantitatea de informație, la limită (vezi proprietatea 4 de mai sus), dacă toți indivizii iau aceeași valoare, aceasta va fi și media, deci indicatorul nostru va reda întreaga informație conținută în caracteristică. Faptul că anumite situații reale sînt departe de acest caz limită nu trebuie să ne ducă la interzicerea calculării mediei, ci la necesitatea de a determina, pe lîngă acest indicator de poziție, și alții, în speță cei de împrăștiere și de asimetrie a distribuțiilor. Dacă însă dorim să reflectăm mai clar valorile mijlocii sau tipice, vom alege dintre indicatorii de poziție pe cei potriviți pentru asta. Unul dintre ei este tocmai cel despre care vom vorbi în continuare.



## 1.2. Mediana

Mediana, notată Me, reprezintă valoarea ce corespunde individului care împarte în două părți egale colectivitatea statistică ai cărei membri sînt presupuși a fi ordonați, în funcție de valorile luate.

O primă precizare se impune imediat. Mediana este o valoare și ea nu trebuie confundată cu individul median sau, cu altă denumire, cu unitatea mediană a populației, reprezentînd individul din populație care ia valoarea mediană. Să exemplificăm definiția cu un exemplu foarte simplu și la îndemînă. Să presupunem că un student la filosofie are într-o sesiune cinci examene, pe care le susține în ordinea următoare și le trece cu notele menționate:

Logică	- 6
Filosofie	- 9
Istorie universală	- 8
Istoria României	- 10
Economie politică	- 6

Ordonînd mulțimea celor 5 unități care formează colectivitatea noastră, de la cea cu valoare minimă la cea cu valoare maximă, obținem:

<u>Unități</u>	<u>Rang</u>
Logică	1-2
Ec.politică	1-2
Ist.universală	3
Filosofie	4
Ist.României	5

Conform definiției și precizări de mai sus, unitatea mediană va fi "Istoria universală" care ocupă rangul mijlociu în ordonarea efectuată, în sensul că înaintea ei sînt două unități și după ea alte două. Valoarea mediană, fiind valoarea individului median, ea este, în cazul nostru, nota 8.





$$Me = 8$$

Prin urmare, pentru determinarea medianei trebuie efectuate următoarele operații: ordonăm indivizii din populație după valorile luate (ceea ce se realizează, de regulă, atunci când se construiește variabila statistică), determinăm rangul individului median și apoi valoarea ce corespunde acestuia. Am văzut că pentru 5 indivizi, rangul unității mediane este 3. În general, pentru  $n$  indivizi,

$$\text{rangul individului median} = \frac{n+1}{2}$$

De pildă, într-o populație cu 999 elemente componente, individul median va avea rangul 500 (1000 : 2)

Se observă că prin formula de mai sus se obține o valoare întreagă numai dacă seria statistică este definită pe un număr impar de indivizi. Într-adevăr, numai atunci există realmente un individ median. În cazul când  $n$  este un număr cu soț, individul median este unul imaginar, al cărui loc se găsește între indivizii de rang  $n/2$  și  $n/2+1$ . De pildă, dacă studentul ale cărui rezultate le-am ilustrat mai sus ar mai fi avut un examen, luat cu nota 10, seria valorilor ar fi:

6, 6, 8, 9, 10, 10

locul individului median imaginar

nu

Dacă există un individ median real, atunci nu există nici o valoare mediană propriu-zisă. Totuși pentru a nu reduce generalitatea aplicabilității indicatorului, se definește valoarea mediană și pentru  $n$  cu soț. Ea este media valorilor pe care le iau indivizii de rang  $n/2$  și  $n/2 + 1$ . În situația de mai sus:

$$Me = \frac{8+9}{2} = 8,50$$

Să vedem, în continuare, cum se rezolvă problema determinării medianei în cazul general, când numărul indivizilor este mare și avem de a face cu caracteristici de genul celor prezentate în capitolul anterior. Vom distinge două situații: când valorile



caracteristicii sînt discrete și efectiv luate de indivizii din fiecare clasă, pe de o parte, și cînd se lucrează cu intervale de grupare, pe de alta.

Pentru primul gen de caracteristici, nu se pune nici o problemă deosebită, rezolvarea fiind identică cu cea din cazurile simple prezentate deja. Dacă numărul indivizilor este fără soț, se va putea identifica rangul unității mediane, care se află într-o clasă caracterizată de o valoare unică, valoarea ce va reprezenta mediana. Dacă numărul unităților este cu soț, se determină rangurile  $n/2$  și  $n/2 + 1$ . Există atunci două posibilități: cei doi indivizi, corespunzînd acestor ranguri, se află în aceeași clasă și atunci mediana este valoarea clasei respective., sau primul se află într-o clasă, caracterizată de valoarea  $x_1$  și al doilea în clasa următoare,  $x_{i+1}$ . În această din urmă situație (foarte improbabilă, de altfel) se va efectua media valorilor celor două clase, ținîndu-se cont de efectivele fiecăreia. Adică:

$$Me = \frac{k_1 x_1 + k_{i+1} x_{i+1}}{k_1 + k_{i+1}}$$

Reluînd exemplul din capitolul anterior cu dimensiunea gospodăriilor (tabelul 6), observăm că mediana trebuie să se găsească între gospodăria cu rangul 125 și 126, deoarece totalul de gospodării este un număr cu soț: 250. Pentru a ajunge la rangurile respective, însumăm succesiv frecvențele:

$$30 + 40 = 70$$

$$70 + 60 = 130$$

Ne oprim aici cîci am depășit cifra 126, și, prin urmare, ambele gospodării căutate se află în clasa cu valoarea 3. Deci:

$$Me = 3$$

sau, altfel spus, gospodăria mijlocie a satului este formată din 3 persoane.

Ca să nu mai construim un exemplu nou, să presupunem, pentru a ilustra cealaltă situație, că mai adăugăm 10 gospodării celor 250, toate posedînd un număr mare de persoane (de pildă,



5 cu 8 membri și alte 5 cu 9). Atunci totalul va ajunge la 260 și mediana se va găsi între cea cu rang 130 și 131. Dar, așa cum am văzut, gospodăria cu rang 130 încheie șirul celor cu dimensiune 3 și deci următoarea va cădea în clasa celor cu dimensiune 4. Cum cele două clase nu au aceleași frecvențe, nu vom putea spune că mediana este 3,50, ci vom lua în considerare numărul de apariții pentru cele două valori. Sau, cu altă exprimare, se efectuează media ponderată a valorilor 3 și 4:

$$Me = \frac{60 \cdot 3 + 50 \cdot 4}{60 + 50} = \frac{380}{110} = 3,45$$

În fine, rămîne să elucidăm situația caracteristicilor realizate prin intervale de grupare. Aici ne vom opri doar la cazul cînd individul median cade într-un interval, deoarece dacă este la limita dintre două intervale - în cazul cînd  $n$  este cu soț, căci altfel individul median e determinat și se va afla sigur într-un interval - , putem fie să atribuim medianei valoarea care desparte cele două intervale, fie să procedăm ca mai sus, utilizînd valorile caracteristice intervalelor. Deci în situația cea mai frecvent întîlnită, individul median se va găsi printre cei care constituie frecvența unui interval de valori, care se va numi interval median. De regulă, interesul cercetătorului este satisfăcut odată cu determinarea intervalului median. De exemplu, privind caracteristica redată în tabelul 7 din capitolul precedent (notele la admitere), se observă imediat că, dat fiind totalul de 600, individul median se află în intervalul de note de la 5 la 5,99, care deci redă nivelul mijlociu al notelor obținute.

Dacă din diferite rațiuni dorim să mergem mai departe și să calculăm o valoare mediană, în situațiile de mai sus, atunci se poate proceda în felul următor. Cînd avem informații detaliate asupra valorilor fiecărui individ, putem urmări situația indivizilor din intervalul median și să-l determinăm pe cel care ocupă poziția de mijloc. De pildă, în cazul mai sus invocat, individul median va fi între cel cu rang 300 și 301. Dar cum pînă la intervalul median sînt 200 indivizi, rămîne doar să găsim, dintre cei 150 cuprinși în intervalul median, pe cel cu rang între 100 (300-200) și 101.



În lipsa informațiilor detaliate despre valorile fiecărui individ din intervalul median, putem proceda - pe baza supoziției că indivizii se distribuie uniform pe acest interval - la un calcul simplu de interpolare. În cazul cu notele, am calculat că pînă la valoarea 5 sînt 200 indivizi și deci pentru a ajunge la mediană trebuie să mai adăugăm 100 (considerăm că individul median este cel cu rang 300, căci oricum calculul este aproximativ). Atunci aplicăm regula de trei simplă:

Dacă la 150 indivizi din intervalul median corespunde o lungime a intervalului de 1 notă, atunci la 100 indivizi va corespunde  $x$ . Si:

$$x = \frac{100 \cdot 1}{150} = 0,67$$

Adăugînd 0,67 la 5, se va obține:

$$Me = 5,67$$

Utilizarea medianei este frecventă și binevenită dacă se urmărește exclusiv indicarea unor mărimi mijlocii, în situații în care distribuțiile sînt puternic asimetrice și deci media este afectată de valori mari sau mici. Din păcate, lipsa unei formule de calcul exprimabilă algebric într-un mod foarte simplu și unitar - ca în cazul mediei - face ca mediana să se găsească departe în urma mediei în întrebuințările statistice. Ea este aproape total evitată atunci cînd trebuie să intre ca element parțial în calculul altor indicatori statistici derivați.

### 1.3. Modul

Modul (și nu modulul, cum greșit mai apare uneori chiar și în manualele de statistică) sau valoarea modală sau dominanta este acea valoare din seria caracteristicii cantitative care este luată de cel mai mare număr de indivizi, deci valoarea căreia îi corespunde frecvența maximă. Valoarea modală se folosește pentru a ilustra cazurile numite tipice în populația studiată, cazurile cele mai frecvente. Este limpede că determinarea acestui indicator



se va face numai pe o masă mare de indivizi, din care se poate degaja o eventuală concentrare a indivizilor pe sau în jurul unei valori date. De pildă, în exemplul cu gospodăriile, se observă că cea mai mare frecvență apare sub valoarea 3, ceea ce înseamnă că în localitatea studiată gospodăria modală (sau tipică) este cea cu 3 persoane.

Determinarea valorii modale se face, după cum se vede, simplu, prin inspecția vizuală a tabelului caracteristicii. Singura problemă apare în cazul intervalelor de grupare când, evident, va trebui în primul moment determinat intervalul modal, adică cel în care frecvența este maximă. Așa de exemplu, în situația candidaților la admitere, intervalul modal al notelor este cel de la 5 la 6. Se poate și aici imagina - ca și pentru mediană - o procedură de calcul a valorii modale. Personal înclin să cred că o astfel de valoare calculată pentru intervale de grupare este prea puțin semnificativă pentru caracteristicile ce redau o situație empirică, deoarece aici efectiv populate sînt intervalele și nu valorile acestora. Desigur, dacă anumite calcule reclamă utilizarea unei valori modale, atunci trebuie utilizate formule pentru determinarea acesteia în cadrul intervalului modal.

## 2. Indicatori de împrăștiere

Indicatorii din această clasă se atasează, de regulă, indicatorilor de poziție pentru a îmbogăți informația asupra caracteristicii studiate. Ei urmăresc, în esență, să măsoare gradul de împrăștiere sau dispersia indivizilor în cadrul scalei de valori pe care le iau. Tradusă problema în alți termeni, se vede că este vorba de a determina gradul de omogenitate sau de eterogenitate al unei populații în raport cu o variabilă statistică. Sau - o altă interpretare posibilă, deși mai puțin uzuală, dar deloc lipsită de importanță - este cea în termeni de inegalitate între indivizi, în raport cu factorul social tradus în caracteristică. O dispersie mare este reflectarea unui grad înalt de inegalitate, pe cînd una mică, specifică populațiilor omogene, ne arată, evident, un grad redus de inegalitate între indivizii populației.

La prima vedere, se pare că problema dispersiei poate fi



ușor expediată urmărind întinderea scalei de valori a caracteristicii. De pildă, dacă în două clase de elevi dăm câte un extemporal la limba română și, în prima clasă, notele obținute vor fi, să zicem, numai între 7 și 9, iar în a doua între 4 și 10, vom putea, bineînțeles, să ne exprimăm că aceasta din urmă este mai eterogenă, că în cadrul ei există diferențe (inegalități) mai mari între elevi decât în cea dintâi. Desigur că și un asemenea indicator de împrăștiere este posibil; el există, poartă denumirea de amplitudine și se notează cu  $A$ . Expresia sa este diferența dintre valoarea maximă și cea minimă în cadrul seriei. Cu notațiile noastre:

$$A = x_8 - x_1$$

Indicatorul amplitudine se poate aplica doar în situații foarte simple, când efectiv ne interesează numai reliefaarea "spațiului de joc" al unei variabile. De cele mai multe ori, el este nesatisfăcător, deoarece face abstracție de frecvențe; într-adevăr problema dispersiei nu se reduce numai la intervalul valorilor luate, ci are în vedere și felul cum se împrăște indivizii pe acest interval. Putem avea două caracteristici identice din punctul de vedere al valorilor, dar foarte diferite din cel al frecvențelor. Spre exemplificare, să recurgem, din nou, la două clase de elevi (imaginare) cărora li se dă un extemporal la limba română. Pentru a ușura citirea datelor, să presupunem că ele sînt de același efectiv; 40 elevi fiecare. Să zicem că distribuțiile notelor sînt cele din tabelul 1.

Note	4	5	6	7	8	9	10	Total
Frecvențe:								
Clasa A	1	2	5	24	5	2	1	40
Clasa B	4	5	6	10	6	5	4	40

Tabelul 1. Distribuțiile notelor în două clase, formate din câte 40 elevi (cifre fictive)

Se observă că, în fiecare clasă, sînt luate toate notele la 4 la 10 de către cel puțin un elev; deci amplitudinile sînt



aceleași:  $10 - 4 = 6$  note. Mai mult, dacă privim cu atenție, vom întui imediat că și mediile notelor pe cele două clase sînt identice: nota 7. Cu toate acestea, există o diferență importantă a modului cum se distribuie elevii noștri. În clasa A, media 7 este rezultatul unui număr foarte mare de note de 7 sau apropiate, pe cînd în B, deși se menține simetria în jurul lui 7, deși această valoare este și aici cea modală, totuși, există un număr mult mai însemnat de indivizi, decît în clasa cealaltă, care se abat de la medie, în sus sau în jos. Putem deci afirma că cele două clase deși se găsesc la același nivel mediu (median sau modal), diferă totuși după gradul de dispersie. În prima avem de a face cu o omogenitate mai mare, în a doua întîlnim un grad superior de dispersie sau de inegalitate: mai mulți copii buni dar tot mai mulți și slabi.

Se pune deci problema calculării gradului de dispersie prin indicatori care să reflecte și situații ca cea de mai sus. În principiu, există două posibilități de a surprinde dispersia indivizilor pe scala valorilor;

- a) fără o valoare de referință față de care să se calculeze abaterile valorilor individuale și
- b) cu o asemenea valoare.

În primul caz, procedura de calcul este destul de complicată, căci trebuie luate în considerare toate cuplurile posibile de indivizi, pentru care se calculează diferența valorilor. Media acestor diferențe luate în valoare absolută ne dă indicatorul de dispersie care poartă numele statisticianului italian Corrado Gini. Formula fiind mai complexă și calculele mai laborioase, indicatorul lui Gini este rareori folosit, mai ales că el nu are avantaje de cunoaștere evitente față de alți indici. În al doilea caz, este posibilă introducerea mai multor indicatori, diversitatea provenind din variantele multiple de alegere a valorii de referință și de calcul al abaterilor față de ea. Voi enumera cîtiva, oprindu-mă puțin asupra celui mai important.

Abaterea medie față de o valoare  $a$ . Este, pur și simplu, media distanțelor dintre valorile  $x_i$  și punctul  $a$ , arbitrar ales, prin distanțe înțelegînd diferențele  $x_i - a$ , în valoare absolută, multiplicată fiecare prin numărul de indivizi care iau valoarea  $x_i$ .



Formula va fi deci:

$$Am(a) = \frac{\sum_{i=1}^n k_i |x_i - a|}{n} \quad (1)$$

Abaterea medie față de medie este un caz particular al indicatorului de mai sus, pentru situația când valoarea fixată, de referință, este chiar media. Se mai numește simplu abatere medie și se calculează cu formula:

$$Am = \frac{\sum_{i=1}^n k_i |x_i - \bar{x}|}{n} \quad (2)$$

Aici se vede mai bine de ce este necesar ca diferențele să fie luate în valoarea absolută. Intr-adevăr, conform unei proprietăți a mediei, suma de la numărător fără valori absolute (suma abaterilor valorilor de la medie) este nulă. Așa însă, diferențele își pierd semnul și nu se mai anulează între ele, rezultatul fiind un număr pozitiv, la limită zero, dacă toate valorile coincid (între ele și deci și cu media lor). Cu cât distanțele valorilor față de medie sînt mai mari și cu cât aceste valori distanțate sînt luate de mai mulți indivizi, cu atît valoarea indicatorului va fi mai ridicată. Indicatorul va arăta, firește, așa cum sugerează și denumirea, cu cât se abate în medie un individ de la valoarea medie a caracteristicii.

După cum se știe, există și o altă modalitate de a înlătura semnul diferențelor decît modulul sau valoarea absolută. Ea constă în ridicarea acestora la pătrat. Procedura ridicării la pătrat este mai frecvent utilizată în statistică datorită algoritmului de calcul mai simplu. Desigur, mărimea astfel obținută este una pătratică și, pentru a se reveni la dimensiunea inițială a variabilei, trebuie extras radicalul. În cazul dispersiei, se vor obține două mărimi similare celor precedente.

Abaterea medie pătratică față de o valoare  $a$ . O vom nota cu  $S(a)$  sau, simplu, dacă nu trebuie precizată valoarea  $a$ ,  $S$ . Formula de calcul va fi



$$S(a) = \sqrt{\frac{\sum_{i=1}^n k_i (x_i - a)^2}{n}} \quad (3)$$

Făcînd  $a$  egal cu media,  $\bar{x}$ , se va obține indicatorul de dispersie cel mai cunoscut și mai utilizat în statistică: abaterea standard ( $\sigma$ ):

$$\sigma = \sqrt{\frac{\sum_{i=1}^n k_i (x_i - \bar{x})^2}{n}} \quad (4)$$

În textele manualelor de statistică sau în alte lucrări din domeniul diferitelor științe, unde se utilizează indicatorul, el poate să mai apară și sub alte denumiri; deviație standard, ecart tip, sau pur și simplu, dispersie. Parțial, această varietate de denumiri în românește se datorează faptului că în engleză indicatorul se numește standard deviation iar în franceză, écart-type. Denumirea de dispersie trebuie oricum înlăturată căci acesta este un termen generic pentru toată problematica indicatorilor respectivi, fiind sinonim cu cel de împrăștiere. Expresia franțuzească de astă dată nu este cea mai ușor de introdus în limba noastră căci termenul de ecart, deși cu oarecare uz, este totuși rareori întâlnit în texte. În expresia englezească, personal înclin - ca majoritatea, cred, a statisticienilor - spre înlocuirea cuvîntului deviație cu cel de abatere, care este consacrat în matematică și statistică pentru a desemna diferențe de genul celor ce intervin aici.

Pătratul abaterii standard ( $\sigma^2$ ) - deci mărimea din formula (4) fără radical - are de asemenea, o largă folosire în statistică, motiv pentru care poartă un nume aparte: varianță. Indicatorul varianță intervine frecvent în analiza relațiilor dintre două sau mai multe variabile statistice. În aceste cazuri, varianța este considerată a fi expresia însăși de manifestare a unui fenomen sau, altfel spus, dar identic din punct de vedere statistic, ca o mărime ce reflectă gradul de nedeterminare a unei variabile. Explicarea unei variabile statistice prin altele este înțeleasă, în mod obișnuit, ca posibilitate de reducere a varianței variabilei



de explicat, atunci cînd variabilele explicative sînt menținute constante sau, în alți termeni, explicarea este reducerea gradului de nedeterminare în raport cu variabila dependentă, dacă se cunoaște poziția individului în raport cu celelalte variabile.

Să trecem acum la ilustrarea procedurilor de calculare a indicatorilor, în speță a abaterii standard. Mai întîi să reluăm exemplul din tabelul 1, pentru a arăta, prin intermediul indicatorilor abatere medie și abatere standard, diferența de dispersie dintre cele două clase de elevi. Vom aranja datele tabelului pe coloane pentru a facilita efectuarea produselor și a însumărilor. Dat fiind că noi cunoaștem de la început media anbelor caracteristici — ea este 7 în cazul de față, calculul din schema ce urmează vizează exclusiv determinarea abaterii standard. De obicei, cum pentru determinarea mediei sînt necesare calcule de genul celor arătate mai înainte, schema de calcul de mai jos se poate lărgi, introducînd ca prime coloane, după cele ale valorilor și frecvențelor, pe cele cerute de calculul mediei.

Pentru clasa A:

$x_i$	$k_i$	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$	$k_i  x_i - \bar{x} $	$k_i (x_i - \bar{x})^2$
4	1	3	9	3	9
5	2	2	4	4	8
6	5	1	1	5	5
7	24	0	0	0	0
8	5	1	1	5	5
9	2	2	4	4	8
10	1	3	9	3	9
Total	40			24	44

Deci

$$Am_A = \frac{24}{40} = 0,60$$

și

$$\sigma_A = \sqrt{\frac{44}{40}} = \sqrt{1,10} = 1,05$$



unde prin indicele A s-a pus în evidență faptul că valorile calculate se referă la clasa școlară (caracteristica) A.

Pentru clasa B:

$x_i$	$k_i$	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$	$k_i  x_i - \bar{x} $	$k_i (x_i - \bar{x})^2$
4	4	3	9	12	36
5	5	2	4	10	20
6	6	1	1	6	6
7	10	0	0	0	0
8	6	1	1	6	6
9	5	2	4	10	20
10	4	3	9	12	36

-----  
Total 40 56 124

Si deci

$$Am_B = \frac{56}{40} = 1,40$$

$$\sigma_B = \sqrt{\frac{124}{40}} = \sqrt{3,10} = 1,76$$

Prin urmare, ambii indicatori atît abaterea medie cît și abaterea standard - pun în relief dispersia mai mare a notelor în clasa școlară B, reflectînd într-o manieră cantitativă ceea ce s-a observat de la început în mod intuitiv. Cum era și firesc, valorile celor doi indicatori, pentru o aceeași caracteristică, nu coincid, căci formulele de calcul sînt diferite. Abaterea medie arată efectiv că, de pildă în clasa B, notele elevilor se abat în medie de la media clasei cu 1,40 puncte. Pentru celălalt indicator, mai clară este interpretarea varianței; la aceeași clasă B, media pătratelor abaterilor notelor individuale de la media clasei este de 3,10 puncte, iar radicalul acesteia este 1,76.

Să ne oprim, în continuare, doar la abaterea standard și să prezentăm o metodă de simplificare a calculelor. În exemplele date, valorile medii au fost din întîmplare (mai exact datorită aranjării intenționate a frecvențelor pentru a ieși așa) cifre ro-



tunde și deci nu a existat nici o dificultate de calcul. În general însă, mediile nu vor mai apărea sub forme atât de simple și deci vom ajunge la calcule mai laborioase atunci când facem diferențele și le ridicăm la pătrat. Pentru ușurarea muncii, ne vom folosi de o proprietate importantă a abaterii standard.

S-a văzut mai sus că abaterea standard reprezintă un caz particular de abatere medie pătratică (când  $a$  coincide cu  $\bar{x}$ ). Fiind dat un  $a$  oarecare și notînd:

$$d = a - \bar{x} \quad (5)$$

atunci se poate scrie o relație simplă între  $S^2(a)$  și  $\sigma^2$ :

$$\sigma^2 = S^2 - d^2 \quad (6)$$

Relația (6) arată că dintre toate abaterile medii pătratice, abaterea standard este cea mai mică. Intr-adevăr, dat fiind că  $d^2$  este o mărime pozitivă sau nulă, din (6) rezultă imediat:

$$\sigma^2 \leq S^2$$

iar inegalitatea se transformă în egalitate atunci când  $d$  este nul, deci când  $a$  coincide cu media.

Formula (6) poate fi folosită pentru a reduce volumul calculelor, deoarece ea ne arată că, înainte de calcularea lui  $\sigma^2$ , putem determina un  $S^2$ , cu o valoare  $a$  convenabil aleasă în acest scop, urmînd ca media să intervină în lucru numai o dată, la sfîrșit, în calculul lui  $d^2$ . Pentru ilustrare, să reluăm exemplul cu dimensiunea gospodăriilor din localitatea rurală. În paragraful precedent am calculat valoarea medie:

$$\bar{x} = 3,66$$

și vom continua acum cu determinarea abaterii standard. Să luăm ca valoare ajutătoare:

$$a = 4$$

Va rezulta

$$d = 4 - 3,66 = 0,34$$



...  $d^2 = 0,1156$

Calcululele ce urmează ne duc la determinarea lui  $S^2(4)$ :

$x_i$	$k_i$	$x_i - a$	$(x_i - a)^2$	$k_i(x_i - a)^2$
1	3	-3	9	27
2	4	-2	4	16
3	6	-1	1	6
4	5	0	0	0
5	3	1	1	3
6	2	2	4	8
7	1	3	9	9
8	5	4	16	80
9	5	5	25	125
Total				895

$$S^2 = \frac{895}{250} = 3,58$$

$$\sigma^2 = 3,58 - 0,1156 = 3,4644$$

$$\sigma = 1,86$$

**Observație:** După cum se poate constata ușor, abaterea standard, ca și ceilalți indicatori introduși aici, se poate scrie și cu frecvențele relative. Intr-adevăr, împărțind fiecare termen al sumei cu  $n$ , formula (4) devine:

$$\sigma = \sqrt{\sum_{i=1}^s p_i (x_i - \bar{x})^2}$$

Din formulele de calcul rezultă că toți coeficienții de dispersie sînt mărimi concrete, în sensul că posedă aceeași unitate



de măsură ca și variabila pe care o reprezintă. Acest lucru facilitează, desigur, înțelegerea semnificației lor dar, în același timp reprezintă și un impediment pentru comparații. Dacă pentru indicatorii de poziție - exprimați și ei în aceleași unități - problema comparației nu se pune decât pentru variabile identice definite pe populații diferite (nu are sens să comparăm, de pildă, vârsta medie cu media numărului de zile lucrate în CAP, ci doar vârsta medie a cooperatorilor dintr-un sat cu cea a cooperatorilor din altul), pentru cei de împrăștiere există situații în care comparațiile apar necesare și pentru variabile diferite definite pe o aceeași populație și chiar - e drept mai rar - pentru variabile diferite rezultate din studiul a două comunități distincte. De exemplu, în cadrul unei colectivități ne putem pune problema dacă inegalitatea indivizilor în raport cu veniturile obținute este mai mare sau mai mică decât în raport cu o altă variabilă cum ar fi consumul cultural (presupunând că s-a putut măsura acesta).

Este deci clar că utilizarea în comparații a abaterii standard este posibilă doar când variabilele sînt aceleași, cum a fost cazul cu notele la extemporale în cele două clase A și B. În alte situații, de genul celei imparate mai sus, va trebui să se calculeze indicatori de dispersie care să nu depindă de unitatea de măsură. Unul dintre aceștia este așa-numitul coeficient de variație  $v$ , dat de formula:

$$v = \frac{\sigma}{\bar{x}}$$

sau, sub formă procentuală:

$$v = \frac{\sigma}{\bar{x}} \cdot 100$$

Prin împărțirea cu media, dispăre unitatea de măsură și indicatorul exprimă proporția sau procentul reprezentat de abaterea standard din medie.



Unele manuale de statistică indică faptul că acest coeficient este expresiv doar atunci când media nu se află în apropierea lui zero, căci altfel valoarea lui este foarte ridicată. Din cele expuse în capitolul referitor la scalele de măsură, se va înțelege de noi că un astfel de indicator se poate utiliza, strict vorbind, doar când scala este de rapoarte, deci când avem de a face cu un zero natural. Intr-adevăr, în cazul scalei de interval, valoarea coeficientului este sensibilă la orice variație a scalei, căci media se modifică în mod corespunzător, pe când abaterea standar rămâne nemodificată. În particular, orice scală de intervale poate fi schimbată astfel ca media să fie nulă și atunci coeficientul de variație nu poate fi calculat.

Importanța indicatorilor de dispersie pentru analiza statistică se poate cu greu intui în întregime după cele spuse mai sus. Abia în capitolele următoare se va întrezări ceva mai clar locul privilegiat pe care aceștia îi ocupă în analizele statistice, îndeosebi în încercările de a modela relațiile dintre fenomene.



## Cap. IX. A S O C I E R E A

### 1. Cazul caracteristicilor dihotomice

Studiul realizat pînă în prezent asupra caracteristicilor sau variabilelor statistice, avînd în vedere tot timpul doar un singur factor în raport cu care se efectuează distribuția indivizilor în clasele (sau după valorile) caracteristicii (variabilei), poate fi considerat, în esență, doar o descriere a populației în funcție de factorul respectiv. Urmează în continuare să încercăm a realiza cîtiva pași pe drumul ce duce la explicarea fenomenelor sociale, utilizînd în acest scop mijloace statistice adecvate.

Orice raționament (iar explicația nu este decît un lanț de raționamente) pornește de la anumite propoziții (judecăți) iar orice judecată presupune două noțiuni care se "leagă" una de cealaltă. Tradus în limbajul nostru - și o să vedem imediat cum se face această traducere - asta înseamnă o analiză a două caracteristici statistice. Dar nu este vorba de a realiza o descriere separată a fiecăreia din ele, ci de o analiză din care să rezulte comportamentul indivizilor simultan în raport cu cele două caracteristici.

Pentru a înțelege mai ușor ce înseamnă această legătură vom începe cu cazul cel mai simplu și anume cel al clasificării dihotomice. O clasificare de acest tip corespunde schemei clasice de analiză a unei noțiuni după sfera și conținutul său. Mai exact, o noțiune definită pe o mulțime de  $n$  indivizi împarte populația noastră în două submulțimi:  $A$  (sau  $A_1$ )<sup>x</sup> - cei care posedă trăsătura definitorie - și  $\bar{A}$  (sau  $A_2$ ) - cei care nu o posedă. Exemplu: fumători/nefumători, religioși/atei etc. Așa cum știm, o asemenea caracteristică se poate reda printr-un tabel de forma:

$A(A_1)$	$\bar{A}(A_2)$	Total
$k_1$	$k_2$	$n$

<sup>x</sup> - Voi folosi pentru clase fie notația  $A$  și  $\bar{A}$ , fie  $A_1$  și  $A_2$ , de la caz la caz.



unde  $k_1$  și  $k_2$  ne arată câți indivizi din cei  $n$  sînt  $A$  (sau  $A_1$ ) și câți sînt  $\bar{A}$  (sau  $A_2$ ).

În mod analog, după un alt atribut  $B$  se realizează o altă clasificare de același tip: o parte dintre indivizi vor intra în clasa  $B$  (sau  $B_1$ ) iar alta în  $\bar{B}$  (sau  $B_2$ ). Dar cum, de regulă, cele două clasificări nu coincid, este clar că în final se vor obține 4 submulțimi în populația celor  $n$  indivizi:

- $A$  și  $B$  ( $A_1$  și  $B_1$ ), cuprinzînd  $k_{11}$  indivizi
- $A$  și  $\bar{B}$  ( $A_1$  și  $B_2$ ), " "  $k_{12}$  " "
- $\bar{A}$  și  $B$  ( $A_2$  și  $B_1$ ) " "  $k_{21}$  " "
- $\bar{A}$  și  $\bar{B}$  ( $A_2$  și  $B_2$ ) " "  $k_{22}$  " "

Prin urmare, din totalul de  $n$  indivizi,  $k_{11}$  se găsesc simultan în  $A$  și  $B$  (sau în  $A_1$  și  $B_1$ ),  $k_{12}$  se află în  $A$  și  $\bar{B}$  etc. Această diviziune se poate pune sub forma unui tabel ca cel de mai jos:

	B		Total
A	$B(B_1)$	$\bar{B}(B_2)$	
$A(A_1)$	$k_{11}$	$k_{12}$	$k_{1.}$
$\bar{A}(A_2)$	$k_{21}$	$k_{22}$	$k_{2.}$
Total	$k_{.1}$	$k_{.2}$	$n$

Tabelul 1.  
Notatii

unde semnificația frecvențelor cu doi indici este cea explicată mai sus. Cît privește frecvențele cu un indice și un punct, care poartă denumirea de frecvențe marginale (cele cu doi indici fiind numite interioare), ele nu redau altoeva decît distribuțiile indivizilor după fiecare caracteristică separat. Rezultă imediat relațiile:

$$k_{11} + k_{12} = k_{1.} \quad k_{21} + k_{22} = k_{2.}$$

$$k_{11} + k_{21} = k_{.1} \quad k_{12} + k_{22} = k_{.2}$$

$$k_{1.} + k_{2.} = k_{.1} + k_{.2} = k_{11} + k_{12} + k_{21} + k_{22} = n$$



Tabelul 1 poartă denumirea de tabel de asociere a caracteristicilor A și B. Se observă din relațiile scrise mai sus că dacă frecvențele marginale sînt date, atunci este posibilă fixarea unei singure frecvențe interioare, celelalte trei rezultă ca diferențe la totalurile de pe liniile sau coloanele tabelului.

Exemplu:

$$k_{12} = k_{1.} - k_{11} \quad \text{etc.}$$

Din acest motiv se spune că tabelul 1 are un singur grad de libertate.

Pentru o mai bună înțelegere a celor expuse pînă aici să dăm un exemplu concret. Să presupunem că într-un studiu asupra unei populații de 100 indivizi am stabilit două caracteristici: A - sexul (cu clasele  $A_1$  - bărbați și  $A_2$  - femei) și B - comportamentul în privința fumatului (cu clasele  $B_1$  - fumători și  $B_2$  - nefumători). Rezultatul - ipotetic - obținut este redat în tabelul 2.

A \ B	B		
	Fumători	Nefumători	Total
Bărb.	30	20	50
Fem.	20	30	50
Total	50	50	100

Tabelul 2.

Avem deci 30 bărbați fumători, 20 bărbați nefumători, 20 femei fumătoare și 30 femei nefumătoare. În total, pe fiecare caracteristică, avem: 50 bărbați și 50 femei; 50 fumători și 50 nefumători. Cu notațiile de mai sus:

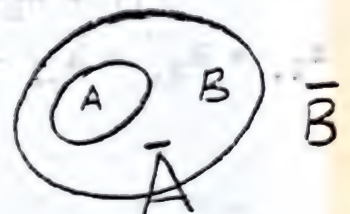
$$k_{11} = 30, \quad k_{12} = 20; \quad k_{1.} = 50 \quad \text{etc.}$$

Pentru a facilita înțelegerea modului de interpretare a unor tabele ca cel de mai sus, voi face apel la una dintre relațiile logice foarte cunoscute: implicația. O implicație de tipul A B poate fi formulată astfel:

"dacă A atunci B"

sau

"toți A sînt B"





cu reprezentarea schematică a sferelor noțiunilor ca mai sus. Consecința unei astfel de implicații, consecință cu repercusiuni asupra formei tabelului nostru, va fi aceea că nu va exista nici un individ care să fie în același timp A și B, prin urmare una dintre căsuțele interioare ale tabelului va conține o frecvență nulă:

$$k_{12} = 0$$

De aici decurge automat că vom avea:

$$k_{11} = k_{1.} \text{ și } k_{22} = k_{.2}$$

Tabelul va fi deci de forma:

	B	B <sub>1</sub>	B <sub>2</sub>	Total
A	k <sub>11</sub>	0	k <sub>1.</sub>	k <sub>1.</sub>
A <sub>2</sub>	k <sub>21</sub>	k <sub>22</sub>	k <sub>2.</sub>	k <sub>2.</sub>
Total	k <sub>.1</sub>	k <sub>.2</sub>	k <sub>.2</sub>	k <sub>.2</sub>

Tabelul 3.

Forma tabelului ce corespunde implicației A → B.

Un tabel de formă asemănătoare se va obține în cazul în care implicația merge de la B spre A. Atunci, pentru că toți B sînt A, nu vom avea nici un individ în intersecția lui B cu  $\bar{A}$  și deci se va anula frecvența  $k_{21}$ . Vor decurge egalitățile:

$$k_{11} = k_{.1} \text{ și } k_{22} = k_{2.}$$

În fine, dacă implicația este reciprocă ( $A \leftrightarrow B$ ), deci toți A sînt B și toți B sînt A, ambele frecvențe de pe diagonala secundară se vor anula:

$$k_{12} = k_{21} = 0$$

și prin urmare:

$$k_{11} = k_{1.} = k_{.1} \text{ și } k_{22} = k_{2.} = k_{.2}$$



Rezultă deci că o implicație reciprocă strictă se poate realiza numai atunci când structurile determinate de cele două caracteristici în populație coincid. În cazul în care mai puțini indivizi posedă atributul A decât B atunci putem avea cel mult o implicație univocă de la A spre B și invers.

Trebuie spus însă că asemenea situații de implicații stricte nu se întâlnesc prea des în realitatea socială și chiar dacă se întâlnesc nu reflectă decât legături banale sau evidente între fenomene. Aceasta este încă o dovadă că faptele sociale se găsesc, așa cum am mai spus, în relații de determinare statistică și nu dinamică. Prin urmare interesul raporturilor descrise mai sus rezidă doar în faptul că ele sînt niște cazuri ideale, față de care realitatea se îndepărtează mai mult sau mai puțin. Dacă dorim totuși să dăm un exemplu concret de implicație strictă am putea imagina situația următoare. Să presupunem că într-o cercetare concretă am surprins comportamentul unui lot de 100 indivizi în raport cu frecventarea bisericii (caracteristica A) și pe de altă parte am reușit să-i delimităm pe cei care cred în religie de cei care nu cred (B). Rezultatul ar putea apare ca în tabelul 4.

Tabelul 4

A \ B	Credincioși Necred. Total		
Merg regulat la biserică	30	0	30
Nu merg regulat la biserică	10	60	70
	40	60	100

Avem deci în cazul tabelului 4 o ipotetică situație de implicație strictă de la A spre B: "Toți cei care merg regulat la biserică sînt credincioși", deși poate că în realitate nici în astfel de cazuri nu obținem frecvența 0 (s-ar putea ca, dintr-un motiv sau altul, în perioada cît efectuăm observația vreun necredincios să frecventeze cu regulatitate biserica).

Situațiile de implicație strictă pot fi interpretate și în altă terminologie. În limbajul teoriei informației, putem apre-



cia că pentru un individ care se află în clasa A (frecventează regulat biserica) deținem o informație completă asupra felului în care se va plasa în clasele lui B: el se plasează cu necesitate în B și nu în  $\bar{B}$ . Sau în limbajul teoriei probabilităților: probabilitatea de a fi B a unui individ care este A este egală cu 1.

Ar mai fi de făcut o precizare referitoare la funcționalitatea relațiilor de implicație strictă în cadrul explicației, mai exact la raportul acestora cu relațiile cauzale. Fără a intra în probleme de epistemologie, precizez doar că, după opinia mea, implicația nu poate fi mai mult decât un indicator care să sugereze o eventuală legătură cauzală. Nu numai că ea nu poate să garanteze o asemenea legătură, dar nici nu-i poate măcar sugera sensul. Iată, chiar în cazul exemplului din tabelul 4, faptul că  $A \rightarrow B$  nu înseamnă că avem o determinare cauzală în același sens. Nu faptul că oamenii merg mai des la biserică determină gradul lor de religiozitate, ci mai degrabă invers. În general, atitudinile și opiniile oamenilor le determină comportamentul și nu invers, deși, evident, între acești doi factori pot exista, cum probabil se întâmplă și în cazul exemplului nostru, interdependențe, reacții de influențare reciprocă.

Concluzia observației precedente este una de o importanță covârșitoare pentru toți cei care utilizează instrumentarul statistic în analiza fenomenelor sociale: numai dintr-o relație statistică, oricât de perfectă ar fi ea, nu se pot trage concluzii asupra fenomenelor reale studiate. Numai introducând ceva din afară, mai bine zis din domeniul teoriei sau al schemelor noastre explicative, constituite înaintea cercetării, de care sîntem sau nu conștienți putem interpreta o legătură statistică.

Dar să mergem mai departe cu analiza. Am spus că în realitate cazurile pure în care sfera unei noțiuni este cuprinsă în sfera alteia - sau, mai mult, cînd ele coincid - sînt rare și neinteresante. Ca urmare ne vor preocupa situațiile în care sferele celor două noțiuni se intersectează. Aceasta logic ar suna: "Unii A sînt B". Statistic însă ne interesează să vedem exact cîți din A sînt și în B, mai bine zis, în ce proporție cei care sînt A sînt și B, în raport cu cei care se găsesc în A. Am putea continua demersul urmînd o cale propusă de R. Boudon prin introducerea noțiunii de implicație slabă, în contrast cu implicațiile strict utilizate mai sus. O



implicație slabă de la A spre B se poate formula: "Dacă A, atunci mai degrabă B decât  $\bar{B}$ ". Tot așa se poate da o expresie și implicației alabe de la B spre A precum și celei reciproce. Un exemplu de implicație slabă de la A spre B este dat în tabelul 5.

	B	$\bar{B}$	Total
A	30	10	40
$\bar{A}$	30	30	60
Total	60	40	100

Tabelul 5.

Se vede că "dacă A, atunci mai degrabă (mai probabil) B decât  $\bar{B}$ ", căci din cei 40 de indivizi din A, 30 se plasează în B și numai 10 în  $\bar{B}$ . Din contră, dacă B, atunci în egală măsură A și  $\bar{A}$  (30 la 30).

În realitate, și cazuri ca acesta sînt rare, în sensul că este puțin probabil ca indivizii dintr-o clasă a unei caracteristici să se găsească exact cîte 50% în clasele celeilalte. În consecință, de cele mai multe ori vom întîlni implicații slabe reciproce, pentru analiza cărora se raționează în felul următor.

Dintre cele două caracteristici alegem una care ar putea determina (cauza, condiționa etc.) pe cealaltă. În cazul în care o asemenea alegere este imposibilă alegem drept bază de plecare oricare din cele două caracteristici. Fie cea aleasă A. Vom încerca apoi să vedem cum se comportă indivizii populației noastre vis-à-vis de cealaltă caracteristică B, în funcție de apartenența lor la una sau alta din clasele lui A. Mai precis, vom calcula proporția indivizilor din A care sînt în B, pe care o vom compara cu cea a indivizilor din  $\bar{A}$  care sînt în B. De exemplu, în cazul tabelului 2, caracteristica de bază este evident sexul, care poate fi un factor ce influențează fumatul. Atunci vom spune:

- dintre bărbați (clasa A sau  $A_1$  - după notațiile noastre) 30 din 50 sînt fumători;
- dintre femei ( $\bar{A}$  sau  $A_2$ ) 20 din 50 fumează;
- în consecință, dintre bărbați fumează mai mulți (sau fumatul este mai intens practicat de bărbați) deoarece raportul 30/50



este mai mare decât 20/50.

În plus, dacă vrem să acordăm o semnificație mai precisă expresiilor "mai mulți" sau "mai intens" este normal că cel mai simplu lucru este să calculăm diferența celor două rapoarte:

$$\frac{30}{50} - \frac{20}{50} = \frac{10}{50} = 0,2$$

diferență care ne dă tocmai o măsură a diferenței de comportament a celor două sexe în privința fumatului.

Poate mai ușor de înțeles stau lucrurile dacă în loc de rapoarte vom lucra cu procente. Calculăm, în cazul de față, procentele de fumători și nefumători în rândul bărbaților, apoi la femeilor și apoi în totalul populației. Se obține tabelul 6.

A \ B	B		Total
	Fumăt.	Nefumăt.	
Bărb.	60	40	100
Fem.	40	60	100
Total	50	50	100

Tabelul 6. Procentele de fumători și nefumători în rândul bărbaților, femeilor și a populației totale studiate, corespunzător tabelului 2.

Se vede acum că dintre bărbații anchetați 60% fumează, dintre femei 40% iar pe ansamblu procentul de fumători este de 50%. Prin urmare, diferența dintre bărbați și femei este de 20%, corespunzând diferenței de rapoarte anterioare de 0,2.

Generalizând modalitatea de analiză de mai sus prin folosirea notației din tabelul 1, vom calcula diferența de comportament a indivizilor din cele două clase ale lui A în raport cu caracteristica B astfel:

$$\delta_A = \frac{k_{11}}{k_{1.}} - \frac{k_{21}}{k_{2.}}$$

În cazul particular al tabelului 2, valoarea lui  $\delta_A$  va fi tocmai 0,2.



În mod analog, se poate calcula un indicator simetric, care să reflecte diferența de comportament a indivizilor din cele două clase ale lui B vis-à-vis de caracteristica A:

$$\delta_B = \frac{k_{11}}{k_{.1}} - \frac{k_{12}}{k_{.2}}$$

În cazul tabelului 2, vom avea:

$$\delta_B = \frac{30}{50} - \frac{20}{50} = 0,2$$

ceea ce înseamnă că cei doi indicatori coincid. Desigur, această coincidență nu este obligatorie, ci ea se produce doar în anumite situații particulare. Se poate demonstra că cei doi indicatori de diferență,  $\delta_A$  și  $\delta_B$ , coincid atunci și numai atunci când se realizează cel puțin una din următoarele trei condiții:

a)  $k_{1.} = k_{.1}$

b)  $k_{1.} + k_{.1} = n$

c)  $k_{11} = k_{1.} \cdot k_{.1} / n$

Condiția a) înseamnă coincidența structurilor determinate în populație de cele două caracteristici, căci ea implică și  $k_{2.} = k_{.2}$ . A doua condiție înseamnă tot coincidența celor două structuri numai că în sens invers; din b) rezultă  $k_{1.} = k_{.2}$  și  $k_{.1} = k_{2.}$ . Despre cea de a treia vom vorbi imediat.

Tot despre cele două valori  $\delta_A$  și  $\delta_B$  se pot demonstra următoarele două proprietăți fundamentale:

1) Cele două diferențe nu pot fi simultan una de un semn și alta de semn contrar.

ii) Cele două diferențe nu pot fi simultan una egală cu și alta diferită de zero.

Pomenind de semnul indicatorilor  $\delta$  am admis deci că ei pot lua și valori negative. Să analizăm un astfel de caz (tab.7) în care cele două caracteristici sînt: A - rata brută de natalitate,



B - PNB pe cap de locuitor, populația statistică fiind constituită din 125 de țări ale lumii.

A \ B	PNB mare	PNB mic	Total
Natal. mare	9	88	97
Natal. mică	26	2	28
Total	35	90	125

Tabelul 7.

De această dată, B apare ca un determinant mai probabil al lui A decât invers. Așadar, în primă instanță, vom calcula

$$\delta_B = \frac{9}{35} - \frac{88}{90} = 0,257 - 0,978 = -0,721$$

care se poate interpreta astfel:

- Țările cu PNB mare au natalitate mare în proporție de 27,7%

- Țările cu PNB mic au natalitate mare în proporție de 97,8%

- Deci avem o diferență de 72,1% de natalitate mare în favoarea țărilor sărace.

Calculînd și indicatorul simetric  $\delta_A$ , se obține:

$$\delta_A = \frac{9}{97} - \frac{26}{28} = 0,093 - 0,929 = -0,835$$

deci tot o valoare negativă, relativ apropiată de prima dar totuși diferită, dată fiind necoincidența celor două structuri marginale.

În sfîrșit, putem trece acum să precizăm și să definim noțiunea de asociere. Se observă că, dacă analizăm simultan comportamentul indivizilor în raport cu cele două caracteristici, plecînd de la o clasă a uneia - să zicem A -, atunci ei au tendința de a se grupa în măsură mai mare într-o clasă a lui B, decât o fac cei care se găsesc în același timp în A. Si invers, plecînd de la B spre A, găsim asemenea tendințe. În general, există două mari posibilități de astfel de grupări: fie cei care sînt în A se plasează



ză preponderent în B (și atunci automat cei din A se grupează în B), fie cei din A se plasează de predilecție în B (și atunci cei din A vor avea tendința să se "îngrămădească" în clasa B). În primul caz vom avea o concentrare pe diagonala principală a tabelului (frecvențele  $k_{11}$  și  $k_{22}$ ) iar în al doilea pe diagonala secundară, corespunzător frecvențelor  $k_{12}$  și  $k_{21}$ .

Această tendință a indivizilor de a nu se distribui uniform în tabel și de a se concentra pe una sau alta din cele două diagonale reprezintă conținutul noțiunii de asociere. De aici și denumirea care se dă în general tabelelor de tipul 1 - tabele de asociere.

Cu ajutorul considerațiilor anterioare, putem defini în mod riguros noțiunea de asociere.

Vom spune că într-un tabel de tipul tabelului 1 există o asociere între caracteristicile A și B dacă valorile  $\delta$  sînt diferite de zero. Vom spune că cele două caracteristici A și B sînt neasociate sau independente dacă valorile  $\delta$  se anulează. (Am menționat că cele două valori  $\delta$  nu pot fi simultan una zero și alta nenulă). Dacă cele două valori  $\delta$  au ambele semnul plus, atunci asocierea se spune că este pozitivă, situație care corespunde concentrării indivizilor pe diagonala principală a tabelului. În caz contrar asocierea este negativă. (Am precizat, de asemenea, că cele două valori  $\delta$ , dacă sînt nenule, trebuie să aibă același semn.)

Referitor la noțiunea de semn sau sens al asocierii se cere făcută o precizare. De multe ori semnul asocierii apare ca ceva artificial, în funcție de felul cum se așează clasele caracteristicilor în tabel. Așa de exemplu, în cazul tabelului 2 vom avea o asociere pozitivă (în sensul precizat că în cei doi  $\delta$  sînt pozitivi - 0,2 - sau că indivizii se concentrează pe diagonala principală). Dacă însă vom scrie întâi linia femeilor și apoi a bărbaților, semnul asocierii se va inversa și el. Ca urmare în astfel de cazuri se cere precizat ce se înțelege prin asociere pozitivă sau negativă: vom zice că există o asociere între sex și fumat, în sensul că bărbații fumează în proporție mai mare decît femeile. Nu același lucru se întîmplă în cazul tabelului 7. Aici este suficient să spunem că între natalitate și produsul național brut pe cap de locuitor există o asociere negativă ca să se înțeleagă faptul că țările cu PNB mare au natalitate mică și invers. În conclu-



zie, putem prezenta ca regulă următoarea: dacă, fixînd clasele unei caracteristici, urmează automat o ordine și pentru clasele celelalte, atunci are sens să vorbim de un semn al asocierii; în caz contrar, nu și trebuie să explicăm direcția asocierii.

După depistarea existenței (inexistenței) și a semnului asocierii (în cazul că există), urmează al treilea pas, care va consta în încercarea de a măsura intensitatea asocierii, respectiv de a construi un indice care să reflecte gradul în care cele două caracteristici sînt legate una de cealaltă.

Există de fapt mai mulți asemenea indici propuși în literatura de specialitate statistică. Unul dintre ei, foarte răspîndit, pe care-l vom numi pur și simplu coeficient de asociere se poate calcula pornind tocmai de la cele două valori  $\delta$ . Coeficientul de asociere, notat cu litera  $\gamma$ , se poate obține luînd media geometrică a lui  $\delta_A$  și  $\delta_B$ , atașîndu-i semnul + sau - în concordanță cu semnul diferențelor  $\delta$ . Deci:

$$\gamma = \pm \sqrt{\delta_A \delta_B} \quad (1)$$

Se poate demonstra ușor că la aceeași valoare se poate ajunge direct din frecvențele tabelului 1. Cu respectivele notații:

$$\gamma = \frac{n_{11} - k_{1.}k_{.1}}{\sqrt{k_{1.}k_{2.}k_{.1}k_{.2}}} \quad (2)$$

formulă în care nu se mai cere atribuirea explicită a unui semn, deoarece acesta rezultă imediat din semnul diferenței de la numărător.

#### Exemple:

1. Să aplicăm formula (2) tabelului 2.

$$\gamma = \frac{100 \times 300 - 50 \times 50}{\sqrt{50 \times 50 \times 50 \times 50}} = \frac{3000 - 2500}{2500} = \frac{500}{2500} = 0,2$$

Intr-adevăr, aici trebuia să se obțină valoarea 0,2 căci, după cum s-a văzut, atât  $\delta_A$  cît și  $\delta_B$  aveau aceeași valoarea 0,2, deci





și media lor geometrică trebuia să iasă tot atât.

2. Pentru situația din tabelul 4:

$$\delta_A = \frac{30}{30} - \frac{10}{70} = 1 - 0,14 = 0,86$$

$$\delta_B = \frac{30}{40} - \frac{0}{60} = 0,75 - 0 = 0,75$$

și

$$\varphi = \sqrt{0,86 \times 0,75} = 0,80$$

sau după forma (2):

$$\varphi = \frac{100 \times 30 - 30 \times 40}{\sqrt{30 \times 70 \times 40 \times 60}} = \frac{3000 - 1200}{100 \sqrt{504}} = \frac{1800}{2245} = 0,80$$

3. Pentru tabelul 7:

$$\varphi = - \sqrt{0,721 \times 0,835} = - 0,776$$

sau

$$\varphi = \frac{9 \times 125 - 97 \times 35}{\sqrt{97 \times 28 \times 35 \times 90}} = - \frac{2270}{2925} = - 0,776$$

#### Proprietăți ale coeficientului de asociere:

1) Limitele de variație a coeficientului  $\varphi$  sînt de la -1 la +1.

ii) Valorile -1 și +1 sînt luate atunci și numai atunci cînd toți indivizii se plasează pe diagonala principală, respectiv, secundară a tabelului, deci în cazurile în care avem implicațiile stricte reciproce  $A \longleftrightarrow B$  sau  $A \longleftrightarrow \bar{B}$ . Am văzut că o condiție necesară pentru a se obține asemenea implicații o constituie coincidența structurilor determinate de A și B în populație, căci toți A trebuie să fie B și reciproc (sau toți A trebuie să fie  $\bar{B}$  și reciproc). Prin urmare, în cazul implicațiilor stricte univoce (numai de la A spre B sau numai de la B spre A), valoarea lui  $\varphi$  nu poate fi 1 sau -1; este cazul situației din tabelul 4, unde  $\varphi = 0,8$ .



iii) Valoarea 0 va fi luată atunci când diferențele sînt nule, deci în cazul de neasociere sau independență a caracteristicilor. Din formula (2) rezultă o condiție necesară și suficientă pentru independență prin anularea numărătorului fracției:

$$nk_{11} - k_{1.}k_{.1} = 0$$

sau

$$k_{11} = \frac{k_{1.}k_{.1}}{n} \quad (3)$$

Se observă că acesta constituie cazul al treilea, c), în care cele două diferențe  $\delta_A$  și  $\delta_B$  coincid (vezi p. 228). Legătura dintre cele trei valori fundamentale ale lui  $\varphi$  (-1, 0, +1) și coincidența lui  $\delta_A$  cu  $\delta_B$  este ușor de înțeles. Datorită faptului că  $\varphi$  este media geometrică a celor două diferențe, el nu poate fi egal cu 1 decât dacă acestea din urmă coincid fiecare cu 1. De asemenea,  $\varphi$  se anulează atunci când diferențele  $\delta$  se anulează. Deci o condiție necesară, dar evident nu suficientă, ca  $\varphi$  să ia valorile critice +1, 0 și -1 este ca valorile  $\delta$  să coincidă.

Mai prezint două noțiuni, legate de fapt de proprietatea ii). Un tabel care redă o situație de implicație strîictă univocă, într-un sens sau altul, vom spune că este un tabel de asociere totală sau maximă, iar unul care prezintă o implicație strîictă reciprocă se va numi de asociere perfectă.

#### Altă metodă de calcul a coeficientului de asociere

Pornind de la definiția asocierii sau independenței putem raționa și în felul următor. Avînd dat tabelul 1, ca rezultat al unei investigații empirice, ne punem problema de a vedea dacă și în ce măsură frecvențele sale interioare diferă de cele ale tabelului care, avînd aceleași distribuții marginale, nu prezintă nici o asociere. Frecvențele unui asemenea tabel, pe care-l numim tabel de independență frecvențe pe care le vom nota cu  $m_{ij}$ , se vor obține cu ajutorul formulei (3):



$$m_{ij} = \frac{k_{i.} \cdot k_{.j}}{n} \quad i, j = 1, 2 \quad (3')$$

Dacă toate frecvențele  $k_{ij}$  coincid cu frecvențele teoretice corespunzătoare  $m_{ij}$ , atunci ne vom găsi în cazul de independență, dacă nu, atunci vom avea o asociere în tabel. Diferențe pot exista numai în două sensuri:

- fie  $k_{11} > m_{11}$ , de unde rezultă automat inegalitățile:  $k_{22} > m_{22}$  și  $k_{12} < m_{12}$  și  $k_{21} < m_{21}$ , deci indivizii se concentrează pe diagonala principală a tabelului,

- fie  $k_{11} < m_{11}$ , de unde rezultă inegalități inverse celor de mai sus, deci indivizii se concentrează pe diagonala secundară.

În felul acesta, prin comparația valorilor  $k$  și  $m$  se poate depista simultan atât existența cât și sensul asocierii. Rămâne problema măsurării ei. Pentru aceasta se face apel la o celebră mărime statistică, notată cu  $\chi^2$  (hi pătrat), dată, în cazul nostru, de formula:

$$\chi^2 = \frac{(k_{11} - m_{11})^2}{m_{11}} + \frac{(k_{12} - m_{12})^2}{m_{12}} + \frac{(k_{21} - m_{21})^2}{m_{21}} + \frac{(k_{22} - m_{22})^2}{m_{22}}$$

sau, mai concentrat

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(k_{ij} - m_{ij})^2}{m_{ij}} \quad (4)$$

Făcînd diferențele căsuță cu căsuță, mărimea  $\chi^2$  măsoară evident abaterea dintre tabelul empiric și cel cu frecvențe întîmplătoare, adică tocmai gradul de asociere din tabel. Dar pentru că  $\chi^2$  depinde de  $n$ , numărul indivizilor din populație, și nu se încadrează într-un interval de variație precis, nu poate fi utilizată ca indice de asociere. Se demonstrează fără dificultate că se poate trece ușor de la  $\chi^2$  la  $\varphi$  prin formula de mai jos:



$$\epsilon = \pm \sqrt{\frac{\chi^2}{n}} \quad (5)$$

semnul atribuindu-se în funcție de situațiile descrise mai înainte.

### Exemple

1. Să presupunem că în urma unei cercetări s-au construit două caracteristici A și B - a căror semnificație nu ne interesează aici - și că asocierea dintre ele se realizează ca în tabelul 8. În tabelul 9 apar frecvențele ce corespund ipotezei de independență.

A \ B	B	B	Total
A	10	40	50
$\bar{A}$	30	20	50
Total	40	60	100

Tabelul 8. Frecvențe empirice

A \ B	B	B	Total
A	20	30	50
$\bar{A}$	20	30	50
Total	40	60	100

Tabelul 9. Frecvențe teoretice

Calculul frecvențelor  $m_{ij}$  din tabelul 9 se realizează după formula (3):

$$m_{11} = \frac{50 \times 40}{100} = 20; \quad m_{12} = \frac{50 \times 60}{100} = 30 \text{ etc.}$$

Vom avea:

$$\begin{aligned} \chi^2 &= \frac{(10-20)^2}{20} + \frac{(40-30)^2}{30} + \frac{(30-20)^2}{20} + \frac{(20-30)^2}{30} = \\ &= \frac{100}{20} + \frac{100}{30} + \frac{100}{20} + \frac{100}{30} = 5 + 3 + 5 + 3,3 = 16,6 \end{aligned}$$

Si

$$\epsilon = - \sqrt{\frac{16,6}{100}} = - 0,41$$

Semnul - a fost acordat în urma observației că în tabelul 8 frecvențele empirice se concentrează pe diagonala secundară.



2. Să reluăm exemplul din tabelul 2. În situația de independență între caracteristicile sex și fumat, s-ar fi obținut:

A \ B	B		
	Fumători	Nefumăt.	Total
Bărbați	25	25	50
Femei	25	25	50
Total	50	50	100

Si deci

$$\chi^2 = \frac{(30-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(30-25)^2}{25} = 1 + 1 + 1 + 1 = 4$$

Si

$$\chi = \sqrt{\frac{4}{100}} = 0,2$$

Observație. Se vede din exemplele de mai sus - și se poate ușor demonstra valabilitatea pe cazul general - că diferențele ce apar la numărătorul fracțiilor în calculul lui  $\chi^2$  sînt egale în valoare absolută. Ele redau surplusul sau deficitul de indivizi în fiecare căsuță a tabelului și este clar că dacă într-o căsuță avem un supliment de indivizi în celelalte două, de pe aceeași linie și coloană, trebuie să avem același deficit.

#### Indicele lui Yule

Un alt indicator folosit pentru a ilustra și măsura gradul de asociere într-un tabel de 2x2 a fost propus de statisticianul englez Yule și este de forma următoare:

$$Q = \frac{k_{11}k_{22} - k_{12}k_{21}}{k_{11}k_{22} + k_{12}k_{21}}$$



Notațiile fiind cele din tabelul 1, se observă că la numărător apare diferența dintre produsul frecvențelor de pe diagonala principală și a celor de pe diagonala secundară, la numitor aceleași produse fiind însumate.

Proprietăți:

- i) Indicele  $Q$  are intervalul de variație cuprins între  $-1$  și  $+1$ .
- ii) Din formulă rezultă imediat oă dacă una din frecvențele  $k_{12}$  sau  $k_{21}$  se anulează,  $Q$  devine egal cu  $+1$ , iar dacă una dintre frecvențele  $k_{11}$  sau  $k_{22}$  este zero,  $Q$  va fi egal cu  $-1$ . În consecință, spre deosebire de  $\chi^2$ ,  $Q$  atinge marginile intervalului de variație pentru implicația strictă univocă.
- iii) Valoarea  $0$  este luată tot în cazul de independență a celor două caracteristici.
- iv) Mărimea lui  $Q$  nu se schimbă dacă înmulțim sau împărțim o linie ori o coloană a tabelului cu un număr. Deci spre deosebire de  $\chi^2$ ,  $Q$  nu este sensibil față de structurile marginale. Această proprietate recomandă folosirea lui  $Q$  în cazul unor studii de eșantion când reprezentativitatea celor două clase ale unei caracteristici nu este asigurată.

Exerciții:

1. Cazul tabelului 2.

$$Q = \frac{30 \times 30 - 20 \times 20}{30 \times 30 + 20 \times 20} = \frac{900 - 400}{900 + 400} = \frac{500}{1300} = 0,38$$

2. Cazul tabelului 4.

$$Q = \frac{30 \times 60 - 0 \times 10}{30 \times 60 + 0 \times 10} = \frac{30 \times 60}{30 \times 60} = 1$$

3. Cazul tabelului 7.

$$Q = \frac{9 \times 2 - 88 \times 26}{9 \times 2 + 88 \times 26} = \frac{18 - 2288}{18 + 2288} = \frac{-2270}{2306} = -0,98$$

4. Cazul tabelului 8.

$$Q = \frac{10 \times 20 - 40 \times 30}{10 \times 20 + 40 \times 30} = \frac{200 - 1200}{200 + 1200} = \frac{-1000}{1400} = -0,71$$



## 2. Cazul clasificărilor multiple

După ce am tratat destul de amănunțit situația particulară a clasificărilor dihotomice, vom trece în continuare să schițăm modalitatea de analiză a caracteristicilor ce divid populația în mai multe clase. Forma generală a tabelului de asociere - sau de contingentă, cum mai poartă denumirea în astfel de cazuri - este cea redată în tabelul 10:

A \ B							
	$B_1$	$B_2$	$B_3$	...	$B_j$	...	$B_t$
$A_1$	$k_{11}$	$k_{12}$	$k_{13}$	...	$k_{1j}$	...	$k_{1t}$
$A_2$	$k_{21}$	$k_{22}$	$k_{23}$	...	$k_{2j}$	...	$k_{2t}$
$A_3$	$k_{31}$	$k_{32}$	$k_{33}$	...	$k_{3j}$	...	$k_{3t}$
...	...	...	...	...	...	...	...
$A_i$	$k_{i1}$	$k_{i2}$	$k_{i3}$	...	$k_{ij}$	...	$k_{it}$
...	...	...	...	...	...	...	...
$A_s$	$k_{s1}$	$k_{s2}$	$k_{s3}$	...	$k_{sj}$	...	$k_{st}$
Total	$k_{.1}$	$k_{.2}$	$k_{.3}$	...	$k_{.j}$	...	$k_{.t}$
	$n$						

Tabelul 10. Forma generală a tabelului de asociere  
(contingentă)

Interpretarea notațiilor din tabelul 10 este următoarea:

$A_1, A_2$  etc. sînt simbolurile celor  $s$  clase ale lui  $A$  iar  $B_1, B_2$  etc. sînt simbolurile celor  $t$  clase ale lui  $B$ ;  $k_{ij}$ , unde  $i$  ia valori de la 1 la  $s$  iar  $j$  de la 1 la  $t$ , sînt frecvențele interioare ale tabelului, respectiv numărul indivizilor aflați simultan în clasa  $A_i$  și  $B_j$ ; în fine,  $k_{1.}, k_{2.}, k_{3.}, \dots, k_{s.}$  și  $k_{.1}, k_{.2}, k_{.3}, \dots, k_{.t}$  sînt frecvențele marginale, adică frecvențele claselor lui  $A$  și  $B$ .





Vom spune că cele două caracteristici A și B sînt independente dacă, oricare ar fi linia i și coloana j, frecvențele empirice  $k_{ij}$  coincid cu frecvențele teoretice  $m_{ij}$  date de formula:

$$m_{ij} = \frac{k_{i.} \cdot k_{.j}}{n}$$

În caz contrar, cele două caracteristici se zice că sînt asociate.

Prin urmare, existența sau inexistența asocierii se poate depista prin compararea tuturor perechilor de valori  $k_{ij}$  și  $m_{ij}$ , perechi care sînt în număr total de sxt. Măsurarea gradului de asociere se va face cu ajutorul unor indici care generalizează indicele  $\varphi$  de la cazul dihotomic, utilizînd mărimea  $\chi^2$ . Din păcate, valoarea  $\varphi$ , dată de formula 5 nu mai posedă în cazul general proprietățile unui indice, în sensul că ea poate depăși unitatea, lucru care nu se întîmplă în cazul dihotomic. Va fi nevoie deci de o ajustare a valorii  $\varphi$  pentru a o aduce la o mărime subunitară. Există mai multe astfel de procedee, dintre care prezentăm două:

- Coeficientul lui Pearson

$$C = \pm \sqrt{\frac{\varphi^2}{1 + \varphi^2}} = \pm \sqrt{\frac{\chi^2}{n + \chi^2}}$$

- Coeficientul lui Ciuprov

$$T = \pm \sqrt{\frac{\varphi^2}{\sqrt{(s-1)(t-1)}}} = \pm \sqrt{\frac{\chi^2}{n \cdot \sqrt{(s-1)(t-1)}}}$$

unde  $\varphi^2 = \frac{\chi^2}{n}$  iar  $\chi^2 = \sum_{i=1}^s \sum_{j=1}^t \frac{(k_{ij} - m_{ij})^2}{m_{ij}}$

Se observă că primul coeficient, C, nu poate atinge niciodată valoarea 1, deoarece numitorul este mai mare decît numărătorul, dar în cazul tabelelor de dimensiune mare el se apropie suficient de mult de unitate. Coeficientul lui Ciuprov ajunge la



valoarea 1, atunci cînd tabelul este pătratic -  $s = t$  - și cînd toți indivizii se plasează pe diagonala tabelului.

Atribuirea semnului plus sau minus coeficienților este o operație mai dificilă acum decît în cazul tabelelor de dimensiune  $2 \times 2$ . În mod riguros, semnul se poate atribui numai atunci cînd sînt îndeplinite simultan următoarele condiții:

- Tabelul este pătratic, deci  $s = t$ .
- Ambele caracteristici sînt construite cu ajutorul unor scale ordinale.
- Dacă clasele unei caracteristici sînt așezate într-o anumită ordine, atunci rezultă logic o ordine și pentru clasele celeilalte.

Dacă cele trei condiții de mai sus sînt îndeplinite, atunci se acordă semnul plus în cazul în care indivizii se plasează pe diagonala principală și semnul minus cînd ei se concentrează pe cea secundară. Dacă prima condiție nu este îndeplinită, tabelul nu va avea căsuțe diagonale. În acest caz se poate totuși atașa un semn ca și mai înainte, lărgind într-o anumită măsură sensul noțiunii de diagonală. Dacă nu sînt satisfăcute condițiile b) și c), atunci o eventuală atașare a semnului trebuie însoțită de o explicație din care să rezulte sensul exact în care se poate vorbi de o asociere.

Dar pentru a înțelege mai bine cele spuse mai sus, precum și felul cum se calculează indicatorii de asociere, să luăm cîteva exemple.

Exercițiul 1. Să presupunem că într-o cercetare de opinie am cerut subiecților să aprecieze în termeni de favorabilitate două evenimente A și B. Datele sînt cele din tabelul 11.

A \ B				Total
	Favorabil	Neutru	Nefavorabil	
Favorabil	20	30	150	200
Neutru	100	300	100	500
Nefavorabil	180	70	50	300
Total	300	400	300	1000

Tabelul 11.



Este limpede că în acest caz ne aflăm în situația când sînt îndeplinite toate cerințele pentru acordarea semnului coeficientului pe care urmează să-l calculăm. Tabelul este de dimensiune 3x3 și dacă am ales pentru A prima clasă "favorabil" este normal s-o alegem și pentru B tot pe aceea. Construim frecvențele întîmplătoare  $m_{ij}$ :

60	80	60
150	200	150
90	120	90

și calculăm:

$$\begin{aligned}\chi^2 &= \frac{(20-60)^2}{60} + \frac{(30-80)^2}{80} + \frac{(150-60)^2}{60} + \frac{(100-150)^2}{150} + \\ &+ \frac{(300-200)^2}{200} + \frac{(100-150)^2}{150} + \frac{(180-90)^2}{90} + \frac{(70-120)^2}{120} + \frac{(50-90)^2}{90} = \\ &= \frac{(-40)^2}{60} + \frac{(-50)^2}{80} + \frac{90^2}{60} + \frac{(-50)^2}{150} + \frac{100^2}{200} + \frac{(-50)^2}{150} + \frac{90^2}{90} + \frac{(-50)^2}{120} + \frac{(-40)^2}{90} = \\ &= \frac{1600}{60} + \frac{2500}{80} + \frac{8100}{60} + \frac{2500}{150} + \frac{10000}{200} + \frac{2500}{150} + \frac{8100}{90} + \frac{2500}{120} + \frac{1600}{90} = \\ &26,7 + 31,2 + 135 + 16,7 + 50 + 16,7 + 90 + 20,8 + 17,8 = 405\end{aligned}$$

Prin urmare:

$$C = -\sqrt{\frac{405}{1000+405}} = -\sqrt{0,288} = -0,54$$

Iar

$$\begin{aligned}T &= -\sqrt{\frac{405}{1000(3-1)(3-1)}} = -\sqrt{\frac{405}{2000}} = \\ &= -0,2025 = -0,45\end{aligned}$$

Semnul minus a fost acordat deoarece avem o concentrare pe diagonala secundară. Acest semn ne spune prin sine că aprecierile celor două evenimente au fost discordante, în general, în sensul că din-



trei cei care au apreciat unul ca favorabil mai mulți au apreciat pe celălalt ca nefavorabil, etc.

Exercițiul 2. Să ne plasăm în aceeași situație ipotetică de mai sus numai să presupunem că pentru A am utilizat o scală cu 5 trepte (tab.12)

A \ B	Favorabil Neutru Nefavorabil Total			
Foarte favor.	0	10	50	60
Favorabil	20	20	100	140
Neutru	100	300	100	500
Nefavorabil	80	50	50	180
Foarte nef.	100	20	0	120
Total	300	400	300	1000

Tabelul 12.

Frecvențele întâmplătoare vor fi:

18	24	18
42	56	42
150	200	150
54	72	54
36	48	36

Calculînd , se obține

$$\chi^2 = 466,8$$

și de aici

$$C = -0,56 \quad ; \quad T = -0,41$$

Semnul minus a putut fi acordat chiar în lipsa unei diagonale propriu-zise, căci se vede că cei care apreciază B ca favorabil apreciază pe A mai degrabă în termeni negativi etc.



**Exercițiul 3.** Să presupunem că asociem o opinie (caracteristica B) cu clasa socială căreia aparține subiectul (A). Se obține tabelul 13.

A \ B				Total
	Favorabil	Neutru	Nefavorabil	
Muncitor	100	200	300	600
Tăran	150	100	50	300
Alte	50	100	50	200
Total	300	400	400	1100

Tabelul 13.

În cazul de față, caracteristica A este abordată cu o scală nominală și prin urmare nu are nici un sens să acordăm asocierii un semn. În astfel de cazuri nu este recomandabil să se treacă direct la calcularea coeficientului de asociere - cum am făcut în cazurile precedente - ci, trebuie analizat tabelul cu ajutorul frecvențelor (procentelor) la totalul unei caracteristici (cea care ar putea-o explica pe cealaltă). În situația noastră, vom calcula procente la totalurile lui A (tabelul 14).

A \ B				Total
	Favorabil	Neutru	Nefavor.	
Muncitor	16,7	33,3	50,0	100
Tăran	50,0	33,3	16,7	100
Alte	25,0	50,0	25,0	100
Total	27,3	36,4	36,4	100

Tabelul 14. Frecvențele tabelului 13 exprimate procentual

Se vede acum foarte bine felul cum se structurează opiniile exprimate prin caracteristica B, în funcție de categoria socială a indivizilor studiați: muncitorii se concentrează spre "nefavorabil", țărani spre "favorabil" iar restul spre "neutru".



Astfel devine limpede că un coeficient de asociere va reflecta aceste "concentrări" de opinii și că este imposibil să-i acordăm un semn. Valorile coeficienților menționați vor fi:

$$C = 0,36 \quad ; \quad T = 0,27$$

Observație. O analiză în termeni de procente, ca cea din exercițiul 3, este recomandabil să fie făcută în toate cazurile deoarece ea ne ajută să vedem foarte clar sensul asocierii mai ales că în situațiile reale nu se obțin cifre atât de rotund ca mai sus și deci inspecția primară a tabelului e mai dificilă.

### 3. Analiza multivariată

Sub titlul de analiză multivariată se înțelege, în domeniul metodologic al sociologiei, ansamblul de proceduri statistice menite să ducă la interpretarea relațiilor complexe ce pot interveni între mai multe caracteristici calitative. Metoda s-a impus aici mai ales în urma remarcabilei lucrări a lui E. Durkheim Le Suicide, a cărei influență s-a făcut simțită cu deosebită vigoare în rândul sociologilor americani. Aceștia au încercat să descopere osatura metodologică a discursului durkheimian, reliefând schemele generale ce pot fi folosite în analiza relațiilor posibile dintre trei sau mai multe caracteristici, scheme care evidențiază anumite cazuri tipice de asemenea relații. Mă voi opri pe scurt asupra celebrei - s-ar putea zice - scheme a lui P. Lazarsfeld, aplicabilă în cazul a trei caracteristici dihotomice (1966).

Să presupunem că ne interesează influența unui factor A asupra altuia B. Primul pas îl constituie, evident, construirea unui tabel simplu de asociere, ca cele folosite în pragrafele precedente, din care se poate deduce existența sau lipsa unei asocieri între A și B. Pasul următor va consta în introducerea unui al treilea factor C, numit, de regulă, variabilă test, despre care se poate presupune că va contribui la elucidarea relației dintre A și B. În esență, e vorba de a scinda populația noastră de n indivizi în două subpopulații, ce constituie efectivele celor două subclase ale lui C: C<sub>1</sub> și C<sub>2</sub>, iar apoi, în cadrul fiecăreia



dintre aceste subpopulații, să se construiască tabele de asociere între A și B. Dacă, de exemplu, în fiecare clasă a lui C, asocierea lui A cu B este aceeași ca și asocierea pe ansamblul populației, se poate trage concluzia că C nu manifestă nici o acțiune asupra relației dintre cei doi factori și, prin urmare, nu aduce nici un plus de informație în analiză. Dacă, dimpotrivă, există diferențe între asocierile din cele două subpopulații - și deci între fiecare subpopulație și ansamblul - , atunci acțiunea lui C trebuie luată în considerare și studiată. Or, tocmai aici și-a adus Lazarsfeld contribuția esențială, evidențiind unele "tipuri ideale" în mulțimea enormă de situații ce pot apare atunci când se asociază simultan trei factori.

Pentru o mai clară precizare a ideilor și pentru a facilita înțelegerea procedurii lui Lazarsfeld, voi începe cu un exemplu concret, construit pe baza unor date fictive, tocmai pentru a ne plasa exact în una din situațiile limită imparate de sociologul american. Fie A factorul sex, cu clasele A<sub>1</sub> pentru bărbați și A<sub>2</sub> pentru femei și B o variabilă care să reprezinte participarea (B<sub>1</sub>) sau neparticiparea (B<sub>2</sub>) la activitățile sportive de masă. Să alegem ca variabilă test, C, faptul că persoanele în cauză efectuează (C<sub>1</sub>) sau nu efectuează (C<sub>2</sub>) munci casnice.

		Efectuează munci casnice ( <u>C</u> <sub>1</sub> )			Nu efectuează munci casnice ( <u>C</u> <sub>2</sub> )			T O T A L		
A	B	Participă ( <u>B</u> <sub>1</sub> )	Nu participă ( <u>B</u> <sub>2</sub> )	Tot.	Participă ( <u>B</u> <sub>1</sub> )	Nu participă ( <u>B</u> <sub>2</sub> )	Tot.	Participă ( <u>B</u> <sub>1</sub> )	Nu participă ( <u>B</u> <sub>2</sub> )	Tot.
Bărbați ( <u>A</u> <sub>1</sub> )		10	90	100	200	200	400	210	290	500
Femei ( <u>A</u> <sub>2</sub> )		40	360	400	50	50	100	90	410	500
Total		50	450	500	250	250	500	300	700	1000

Tabelul 15. Distribuția persoanelor după cele trei caracteristici, A, B și C.



Tabelul 15 prezintă datele ipotetice obținute. Se observă din ultima treime a tabelului că, pe ansamblul celor 1.000 de persoane investigate, există o asociere strînsă între sex și participarea la acțiunile sportive. Bărbații iau parte în proporție de 42% (210 din cei 500), pe cînd femeile numai de 18% (90 din cele 500). Privind acum primele două părți ale tabelului, în care sînt separate cele două loturi după C - 500 persoane care efectuează munci casnice și 500 care nu efectuează -, vedem imediat că diferența pe sexe dispare: în primul caz, atît participarea bărbaților cît și a femeilor este de 10% (10 din 100, respectiv 40 din 400), iar în al doilea de 50% (200 din 400 și, respectiv, 50 din 100).

Din datele prezentate (cu repetarea precizării că ele sînt absolut fictive), rezultă că numai antrenarea în munci casnice, diferită pentru cele două sexe, este responsabilă de inegala participare a bărbaților și femeilor la acțiunile sportive. Dacă am fi analizat doar asocierile simple între cei trei factori (A cu B, A cu C, B cu C), s-ar fi observat că bărbații sînt mai puțin prinși în munci casnice, apoi că cei ce efectuează asemenea munci au o participare mai redusă la activitatea sportivă, ceea ce implică imediat asocierea existentă între sex și practicarea sportului. Dar această implicație este una slabă și ea nu este capabilă să ne indice intensitatea asocierii lui A cu B și, cu atît mai puțin, să ne conducă la concluzia că întreaga asociere a lui A cu B se datorează variabilei test C. Numai analiza tabelului de asociere simultană a celor trei factori poate pune cu claritate în evidență această concluzie.

Dacă notăm cu  $\chi(A,B)$ ,  $\chi(A,C)$ ,  $\chi(B,C)$  coeficienții de asociere între cele trei perechi de caracteristici și cu  $\chi(A,B;C_1)$  și  $\chi(A,B;C_2)$  asocierea dintre A și B în cele două clase ale lui C, deci coeficienții parțiali de asociere ai lui A cu B cînd C este menținut constant, atunci se poate scrie o egalitate fundamentală, datorată lui Lazarsfeld:

$$\chi(A,B) = \alpha \cdot \chi(A,B;C_1) + \beta \cdot \chi(A,B;C_2) + \gamma \cdot \chi(A,C) \cdot \chi(B,C)$$

unde  $\alpha$ ,  $\beta$  și  $\gamma$  sînt trei coeficienți pozitivi.



În exemplul prezentat mai sus, se vede că vom avea:

$$\chi^2(A, B; C_1) = \chi^2(A, B; C_2) = 0$$

deci relația lui Lazarsfeld va avea forma:

$$\chi^2(A, B) = 0 + 0 + r \cdot \chi^2(A, C) \cdot \chi^2(B, C)$$

care caracterizează tipul numit de sociologul american "elaborare prin relații marginale (M)" iar de către R. Boudon (1967) "structură cu efect de regresie". Acest tip se descompune, la rândul său, în alte două subtipuri, după cum variabila test este anterioară lui A, cazul MA sau intermediară lui A și B, cazul MI.

Distincția dintre situațiile MA și MI nu este deloc convențională, ci are implicații serioase în puterea explicativă a modelului celor trei variabile analizate. Într-adevăr, se vede că, în general, într-o situație de tipul M, relația statistică a lui A cu B se datorează în exclusivitate factorului C. Dacă C se plasează între A și B, cum se întâmpla în exemplul nostru, atunci putem spune că A nu are o influență directă asupra lui B și că influența sa se transmite prin intermediul lui C. Influența sexului asupra participării la activitatea sportivă se materializează în diferența de participare la activitățile casnice. Analog este și exemplul pe care-l dă Lazarsfeld: diferența dintre tineri și vîrstnici în audiența la radio a emisiunilor religioase - în favoarea celor din urmă - se explică prin intervenția factorului instrucție, inegal distribuit în cele două grupe de vîrstă, intervenție care se concretizează prin faptul că la fiecare nivel de instrucție asocierea dintre vîrstă și ascultarea emisiunilor religioase dispare.

Schematic, situația MI se poate reprezenta astfel:



În cazul în care variabila test este anterioară factorului explicativ A, asocierea statistică dintre A și B este absolut străină unei relații de influențare, atunci cînd sînt satisfăcute condițiile tipului M (anularea asocierilor parțiale). Ea se datorează și acum factorului test, C, numai că el fiind anterior și lui A și lui B, nu mai e vorba de o legătură propriu-zisă, nici

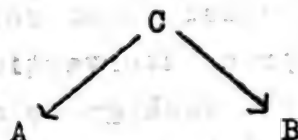


măcar mediată, între A și B. Dacă transpunem în limbaj cauzal situația MA, atunci putem spune că relația lui A cu B rezultă din faptul că ambele fenomene sînt efectele lui C. Exemple pentru această situație se pot imagina destule, dar prefer să reproduc două care au devenit deja clasice în manualele de statistică:

a) Se constată, pentru o mulțime de localități, o asociere strînsă pozitivă între numărul de berze din localitate (factorul A) și rata de natalitate în localitatea respectivă (B). Legătura dintre A și B este absolut iluzorie, fapt care se poate demonstra dacă se introduce ca variabilă test (C) tipul de localitate (oraș-sat). În mediul rural, berzele sînt mai numeroase decît la orașe iar natalitatea este de asemenea, mai mare.

b) Se constată, la fel, o asociere, foarte strînsă între cantitatea de apă aruncată de pompieri în cursul unui incendiu (factorul A) și gravitatea pagubelor produse de incendiu (B). Evident că nici această asociere nu reflectă o relație de determinare de la A spre B; un al treilea factor, anterior lui A și B, pe care-l putem denumi "gravitatea incendiului" explică atît pagubele produse cît și volumul apei utilizată de pompieri.

Pentru situația MA, schema grafică este următoarea:



Al doilea tip principal prezentat de Lazarsfeld este cel în care asocierea dintre A, variabila independentă, și C variabila test, este nulă. El numește acest caz "elaborare prin relații parțiale (P)" iar Boudon preferă expresia de "structură cu efect de interacțiune". Prima denumire este justificată de faptul că, datorită anulării lui  $\chi^2(A, C)$ , ecuația lui Lazarsfeld devine:

$$\chi^2(A, B) = \alpha \cdot \chi^2(A, B; C_1) + \beta \cdot \chi^2(A, B; C_2)$$

adică asocierea lui A cu B apare doar ca o resultantă a asocierilor parțiale:  $\chi^2(A, B; C_1)$  și  $\chi^2(A, B; C_2)$ . A doua subliniază faptul că tocmai aceste asocieri parțiale sînt indicatorii de măsurare a efectului interacțiunii lui A și C în determinarea lui B.



Să luăm și pentru cazul de față un exemplu, de asemenea cu date fictive, reprezentat în tabelul 16.

		Instrucție su- perioară ( $C_1$ )			Instrucție infe- rioară ( $C_2$ )			T O T A L		
A	B	Inte- res ( $B_1$ )	Dezin- teres ( $B_2$ )	Tot. ( $B$ )	Inte- res ( $B_1$ )	Dezin- teres ( $B_2$ )	Tot. ( $B$ )	Inte- res ( $B_1$ )	Dezin- teres ( $B_2$ )	Tot. ( $B$ )
Bărba- ți ( $A_1$ )		130	20	150	250	100	350	380	120	500
Femei ( $A_2$ )		120	30	150	100	250	350	220	280	500
Total		250	50	300	350	350	700	600	400	1000

Tabelul 16. Un exemplu de elaborare prin relații parțiale

Fie A variabila sex, B - interesul pentru probleme de politică externă și variabila test, C - nivelul de instrucție școlară. Cifrele din tabelul 16 arată că sexul nu este asociat cu instrucția: 150 de bărbați din cei 500 și 150 de femei tot din 500 au nivelul superior de instrucție. Dimpotrivă, atât sexul cât și instrucția se află într-o relație statistică destul de strînsă cu interesul pentru politica externă: 380 de bărbați se interesează, față de numai 220 de femei, din același efectiv total de 500; 5/6 (250 din 300) dintre cei cu instrucție superioară, față de numai jumătate (350 din 700) dintre cei cu nivel școlar scăzut. În atare condiții este deosebit de importantă informația adusă de asocierile parțiale. Este imediat vizibil că în subșantionul celor cu instrucție superioară (clasa  $C_1$ ) asocierea dintre sex și interes este deosebit de modestă, aproape nulă: bărbații se interesează în procent de 86,7% (130 din 150) iar femeile în proporție de 80% (120 din 150). Din contră, în celălalt subgrup ( $C_2$ ) proporția bărbaților interesați de politică depășește spectaculos pe cea a femeilor: 71,4 % (250 din 350) față de numai 28,6% (100 din 350). Prin urmare, dacă datele prezentate ar fi reale,



am putea afirma că influența sexului asupra interesului pentru probleme politice se materializează doar în interacțiune cu educația școlară, în sensul că diferențele dintre sexe se șterg la nivelurile superioare de instrucție, rămân clare doar la cele inferioare.

Exemplul acesta poate fi interpretat ca reflectând o situație experimentală cu grup de control. Fenomenul A apare drept stimul, C drept dispoziție iar B răspuns. Într-un astfel de limbaj se înțelege mai ușor de ce diferența de asociere dintre A și B în cele două clase ale lui C poate fi numită efect al interacțiunii lui A cu C. (Stimulul interacționează cu dispoziția în producerea răspunsului).

Există și de data aceasta două subcazuri. Unul, ca cel reflectat de exemplul de mai sus. când C este intermediar lui A și B, este notat de Lazarsfeld cu simbolul PI; celălalt, notat cu PA, corespunde situației în care C este anterior lui A. Distincția dintre PI și PA nu este atât de importantă ca cea dintre MI și MA, în cele două cazuri P, schimbându-se doar rolurile variabilei test și a celei independente.

Interpretarea în limbajul experimentului se datorează aici faptului că asocierea dintre A și C este nulă. Într-adevăr dacă cei care posedă calitatea A<sub>1</sub> constituie grupul experimental iar cei care n-o posedă (cei din A<sub>2</sub>), grupul de control, lipsa de asociere dintre A și C nu înseamnă altceva decât că, din punctul de vedere al comportamentului față de C, grupurile experimentale și de control sînt identice. Or, aceasta era una dintre supozițiile fundamentale ale experimentului cu grup de control. Dacă situația reală nu ne conduce la un astfel de raport între A și C, ea poate fi "forțată" să se încadreze într-o asemenea schemă, alegînd convenabil indivizii din A<sub>2</sub> sau din A<sub>1</sub>. În cazul nostru, de pildă, dacă bărbații și femeile n-ar fi avut aceeași distribuție după școlaritate, s-ar fi putut alege un lot de femei (sau de bărbați) care să aibă o structură școlară identică cu a bărbaților (sau femeilor). Desigur că atunci corelațiile ansamblului eșantionului  $\chi(A,B)$ ,  $\chi(A,C)$  și  $\chi(B,C)$  ar fi fost denaturate într-o oarecare măsură, dar, la urma urmei, ele nu prezintă interes ca valori exacte. Diferența de comportament a indivizilor din cele două clase ale lui A în raport cu variabila B s-ar menține (putînd fi măsu-



rată prin  $\delta_A$ ) iar interacțiunea dintre A și C în determinarea lui B ar ieși mai clar în evidență prin anularea lui  $\psi(A,C)$

În general, se cuvine observat că situațiile reale rareori se suprapun peste vreunul din tipurile ideale descrise, sau pot fi aduse la acestea. De cele mai multe ori, sîntem conduși să analizăm tabele fără a putea măcar spunecoă ele se apropie de tipurile lui Lazarsfeld, ceea ce nu diminuează cu nimic importanța distincțiilor teoretice de mai sus. Într-adevăr, din analiza precedentă rezultă unele învățăminte de care trebuie să ținem seama atunci cînd analizăm tabele empirice efectiv rezultate dintr-o cercetare:

a) Dacă vom găsi că mărimile asocierilor parțiale

$\psi(A,B;C_1)$  și  $\psi(A,B;C_2)$  sînt aproximativ egale și au același semn cu asocierea de ansamblu  $\psi(A,B)$ , atunci înseamnă că o parte din asocierea lui A și B trebuie pusă pe seama lui C, iar această parte este cu atît mai mare cu cît asocierea totală diferă mai mult de asocierile parțiale. În limită, putem ajunge în două situații: cînd asocierile parțiale sînt nule (cazul M) - și deci întreaga corelație (A,B) se poate pune pe seama lui C - sau cînd asocierile parțiale coincid cu asocierea totală perfect:

$$\psi(A,B;C_1) = \psi(A,B;C_2) = \psi(A,B)$$

și atunci introducerea factorului C în analiză este superfluă, deoarece acesta nu afectează cu nimic raportul lui A cu B. În toate situațiile, poziția lui C (înainte de A sau între A și B) este foarte importantă pentru interpretarea relațiilor dintre variabile.

b) În cazul în care asocierile parțiale diferă semnificativ una de cealaltă, atunci, indiferent de faptul că  $\psi(A,C)$  este sau nu egal cu zero, înseamnă că că ne găsim în situația de interacțiune. Cu cît diferența este mai mare cu atît interacțiunile sînt mai puternice și nu este exclus să întîlnim situațiile reale în care semnele celor două asocieri parțiale să fie contrare. Ne-am fi găsit într-o asemenea situație dacă, de pildă, în exemplul precedent, la nivelul de instrucție superior, femeile ar fi fost cele care se interesează mai mult de politică, iar la cel inferior, bărbații.



Este evident că, dacă cele două asocieri parțiale sînt de semne contrare, se poate ajunge - prin "raportul de forțe" al ponderilor diferitelor clase ale variabilelor - la situația în care asocierea totală  $(A, B)$  să fie nulă, în condiții deci de influență a lui A asupra lui B. Prin urmare, nici existența unei asocieri între A și B nu reprezintă o garanție a unei legături între factorii respectivi (vezi cazul MA) și nici lipsa unei asocieri nu este o dovadă a independenței reale a celor doi factori.

Am pomenit mai înainte despre apropierea dintre tipul P și o situație experimentală; nu este posibil să nu se remarce faptul că de fapt orice analiză multivariată încearcă o rafinare a datelor de observație în scopul apropierii de condițiile experimentului. Este evident că introducerea variabilei test și analiza relațiilor dintre factorul "independent" și cel "dependent", în clasele variabilei test, nu înseamnă altceva decît controlul lui C. Dar, după cum am spus cînd am vorbit despre experiment, controlul trebuie să se aplice tuturor factorilor care pot influența raportul celor două variabile de bază. Această idee ne conduce la extinderea procedurii analizei multivariate, extindere ce se poate realiza pe două căi: fie prin introducerea succesivă a mai multor variabile test, fie prin introducerea simultană a acestora. Desigur, că din punctul de vedere al rezultatelor de cunoaștere, este preferabilă această din urmă variantă, deoarece ea realizează mai fidel una din condițiile experimentale (controlul simultan al factorilor), dar, din păcate, practic o astfel de analiză este greu de realizat. În adevăr, chiar și pentru cazul cel mai simplu al caracteristicilor dihotomice, numărul căsuțelor interioare - deci fără a mai socoti toate totalizările parțiale ce se pot face - crește foarte repede: pentru 2 variabile test - 16 căsuțe, pentru 3 - 32 căsuțe, pentru 4 - 64 etc. Cînd clasificarea nu este dihotomică, cifrele de mai sus devin amețitor de mari. Presupunînd, de pildă, că fiecare variabilă are cinci clase, un tabel cu o variabilă test va avea 125 de căsuțe interioare, unul cu 2 variabile test - 625 de căsuțe, cu 3 variabile test - 3125 ș. a.m.d. Creșterea numărului de căsuțe în tabel îngreunează analiza ducînd-o pînă la urmă la imposibilitate, din două motive. Mai în-  
tîi, este clar că nici o minte, oricît de versată ar fi în analize



statistice, nu poate "vedea" clar ce se petrece într-un tabel de asemenea dimensiuni. Apoi, și ceea ce este mai important, aceste tabele "mamut" trebuie să conțină cifre în fiecare căsuță iar frecvențele respective trebuie să fie suficient de mari pentru a putea face comparații valide între diferitele rubrici ale tabelului. De aici rezultă necesitatea unor eșantioane enorme, de care sociologul rareori dispune.

Voi prezenta, în încheiere, un exemplu de analiză multivariată cu cifre reale, exemplu ce mai are și particularitatea că numărul claselor este variabil în cele trei caracteristici. La un examen de admitere în învățământul superior - vezi T. Rotariu (1974) - s-a urmărit, printre altele, construcția caracteristicilor de mai jos:

- Originea socială a candidaților (A), distingându-se trei categorii:

- $A_1$  - intelectuali
- $A_2$  - categorii intermediare
- $A_3$  - muncitori și țărani

- Pregătirea școlară din liceu (C), formându-se patru clase, în funcție de notele de la bacalaureat și din ultima clasă de liceu:

- $C_1$  - medii între 9 și 10
- $C_2$  - " " 8 și 8,99
- $C_3$  - " " 7 și 7,99
- $C_4$  - " sub 7

- Reușita la admitere (B) variabila dependentă, cu două clase:

- $B_1$  - admis
- $B_2$  - respins



Pregătire școlară (C)	C <sub>1</sub>			C <sub>2</sub>			C <sub>3</sub>			C <sub>4</sub>		
Originea (A)	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>
% de reu- șiți (B)	70,4	63,0	57,4	54,5	42,6	37,8	31,1	25,9	21,4	20,7	16,0	15,8

**Tabelul 17. Influența pregătirii școlare și originii sociale asupra reușitei candidaților la admiterea în învățământul superior.**

Tabelul 17 concentrează întreaga informație necesară analizei, cu excepția corelațiilor globale, care sînt toate pozitive. Plecînd de la asocierea dintre A și B, în sensul că fiii de intelectuali reușesc în procent mai ridicat, s-a pus problema dacă factorul intermediar C transmite sau nu toată influența lui A asupra lui B. Datele noastre ne arată că nu ne găsim în cazul de elaborare prin relații marginale, MI, deoarece la fiecare nivel de pregătire școlară se menține o asociere între origine și reușită, prin urmare există o cale de influențare ce nu trece prin C.

A doua constatare ce se impune din analiza datelor tabelului este existența unei interacțiuni între originea socială și pregătirea școlară în influențarea reușitei la admitere: pe măsură ce coborîm pe scara notelor - de la C<sub>1</sub> la C<sub>4</sub> - , efectele directe ale originii sociale se diminuează.



## Cap. X. CORRELATIA

Problematica ce constituie conținutul acestui capitol este similară celei expuse în cazul asocierii, cu deosebirea că, de această dată, ne vom referi la caracteristici cantitative. Prin urmare, noțiunea de corelație înlocuiește, pentru variabilele statistice, pe cea de asociere, folosită spre a ilustra interdependența caracteristicilor calitative.

S-ar părea că vorbind aici despre corelație ar trebui să repetăm în mare cele spuse cu ocazia prezentării asocierii, adăugând ceva în plus din cauza proprietăților specifice variabilelor (clasele acestora sînt caracterizate de valori numerice). S-ar părea deci că, în esență, corelația nu este altceva decît asocierea aplicată la variabile. Lucrurile nu stau însă deloc așa; este drept că între cele două noțiuni există o înrudire foarte strînsă dar ele nu trebuie totuși confundate sau, mai corect spus, corelația nu este o speță (un caz particular) de asociere. Este evident că, din faptul că variabilele pot fi privite drept caracteristici calitative, făcînd abstracție de valori, rezultă imediat că tot ce s-a vorbit în capitolul precedent despre acestea din urmă este valabil și aplicabil variabilelor. Adică noțiunea de asociere - respectiv independență - se poate folosi și la variabile fără nici o restricție.

Calea prin care am ajuns la definirea asocierii nu se petrece însă la dezvoltări fructuoase pentru variabile - datorită faptului că nu permite introducerea în calcule a valorilor - și de aceea o vom părăsi, alegînd o altă care ne va conduce la definirea corelației. Se va reveni apoi spre a compara cele două noțiuni și a ilustra succint asemănările și deosebirile dintre ele.

### 1. Convarianță, coeficient de corelație

Fie  $X$  și  $Y$  două variabile ce iau valorile  $x_1, x_2, \dots, x_s$  și, respectiv,  $y_1, y_2, \dots, y_t$ , valori scrise în ordinea lor crescătoare, Să notăm cu  $k_1, k_2, \dots, k_s$  și cu  $k_{.1}, k_{.2}, \dots, k_{.t}$  frecvențele cu care sînt luate aceste valori iar cu  $k_{ij}$  ( $i=1, 2, \dots, s$ ;  $j=1, 2, \dots, t$ ) numărul indivizilor care iau simultan valorile  $x_i$  și  $y_j$ .



Forma generală a tabelului de corelație (tabelul 1) este similară celei a tabelului de asociere (tabelul 10, cap.9)

Să luăm un individ oarecare din populația noastră și să presupunem că, referitor la variabila X, el primește valoarea  $x_1$ , iar, referitor la Y, pe  $y_j$ . Dacă  $\bar{x}$  și  $\bar{y}$  sînt valorile medii ale celor două variabile, putem construi diferențele:

$$x_1 - \bar{x} \text{ și } y_j - \bar{y}$$

$\begin{matrix} Y \\ \backslash \\ X \end{matrix}$	$y_1$	$y_2$	$\dots$	$y_j$	$\dots$	$y_t$	Total
$x_1$	$k_{11}$	$k_{12}$	$\dots$	$k_{1j}$	$\dots$	$k_{1t}$	$k_{1.}$
$x_2$	$k_{21}$	$k_{22}$	$\dots$	$k_{2j}$	$\dots$	$k_{2t}$	$k_{2.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$x_i$	$k_{i1}$	$k_{i2}$	$\dots$	$k_{ij}$	$\dots$	$k_{it}$	$k_{i.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$x_s$	$k_{s1}$	$k_{s2}$	$\dots$	$k_{sj}$	$\dots$	$k_{st}$	$k_{s.}$
Total	$k_{.1}$	$k_{.2}$	$\dots$	$k_{.j}$	$\dots$	$k_{.t}$	$n$

**Tabelul 1. Forma generală a tabelului de corelație**

Dacă una dintre diferențe este pozitivă, atunci individul în cauză se află plasat, prin valoarea luată, în dreapta mediei și, dacă e negativă, în stînga. De exemplu:

$$\begin{array}{ccccccc}
 x_1 & x_2 & \dots & x_1 & \dots & \bar{x} & \dots & x_n \\
 & & & \uparrow & & \uparrow & & \\
 & & & \text{diferență ne-} & & & & \\
 & & & \text{gativă} & & & & 
 \end{array}$$

În raport cu ambele variabile, pentru un același individ putem avea patru situații:



$$a) \begin{cases} x_1 - \bar{x} \geq 0 \\ y_1 - \bar{y} \geq 0 \end{cases}$$

$$b) \begin{cases} x_1 - \bar{x} \leq 0 \\ y_1 - \bar{y} \leq 0 \end{cases}$$

$$c) \begin{cases} x_1 - \bar{x} \geq 0 \\ y_1 - \bar{y} \leq 0 \end{cases}$$

$$d) \begin{cases} x_1 - \bar{x} \leq 0 \\ y_1 - \bar{y} \geq 0 \end{cases}$$

Primele două ne arată o concordanță în plasarea indivizilor pe scala valorilor lui X și Y, în raport cu mediile acestora, pe cînd ultimele o discordanță. Cele patru cazuri se pot, în aceeași idee, sintetiza în două. Intr-adevăr, prin efectuarea produselor diferențelor, în cazurile a) și b) se va obține o mărime pozitivă (eventual nulă), iar pentru c) și d) una negativă (sau nulă):

$$\alpha) (x_1 - \bar{x})(y_1 - \bar{y}) \geq 0 \quad - \text{concordanță}$$

$$\beta) (x_1 - \bar{x})(y_1 - \bar{y}) \leq 0 \quad - \text{discordanță}$$

Să ne imaginăm, în continuare, că vom efectua aceste produse pentru fiecare individ din populație, încercînd să vedem care situație este predominantă: cea de concordanță sau cea de discordanță. Desigur că nu este suficient să numărăm cazurile de un tip și de celălalt și să facem bilanțul, dat fiind că aceste produse nu au numai un semn dar și o valoare absolută, care trebuie luate în considerare. Se pune deci problema de a găsi o valoare unică destinată să caracterizeze, într-o manieră sintetică, felul cum se "comportă" ansamblul indivizilor simultan în raport cu cele două variabile. Adică să arate în ce măsură (și termenul de măsură este folosit aici în sensul său propriu) indivizii care iau valori mari (mici) pentru X iau valori mari (mici) și pentru Y. Or, un indicator potrivit pentru o atare măsură este media aritmetică a produselor efectuate pentru toți indivizii.

Privind tabelul de corelație, vedem că nu este nevoie să facem pe asemenea produse. De pildă,

$$(x_1 - \bar{x})(y_1 - \bar{y})$$

este comun mai multor indivizi, și anume pentru toți cei aflați în căsuța din colțul stîng și de sus al tabelului, în număr de  $k_{11}$ .



Suma acestui produs pentru toți cei  $k_{11}$  indivizi va fi evident:

$$k_{11}(x_1 - \bar{x})(y_1 - \bar{y})$$

În general, vom ajunge, pentru calcularea mediei, la o sumă de termeni de forma:

$$k_{ij}(x_i - \bar{x})(y_j - \bar{y})$$

unde indicele  $i$  ia valori de la 1 la  $s$ , iar indicele  $j$  de la 1 la  $t$ , deci la o sumă de  $s \cdot t$  termeni (numărul liniilor înmulțit cu numărul coloanelor sau numărul căsuțelor interioare ale tabelului), sumă care finalmente trebuie împărțită la  $n$ , numărul total de indivizi. Mărimea astfel obținută se numește covarianța variabilelor  $X$  și  $Y$  și se notează cu  $\text{Cov}(X, Y)$ . Prescurtat, ea se scrie:

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^s \sum_{j=1}^t k_{ij}(x_i - \bar{x})(y_j - \bar{y})}{n}$$

În cazul în care  $\text{Cov}(X, Y)$  este nulă, vom spune că variabilele  $X$  și  $Y$  sînt necorelate, iar dacă este diferită de zero, că ele sînt corelate. Cînd  $\text{Cov}(X, Y) < 0$ , vom avea o corelație negativă (însemnînd că discordanțele de comportament prevalează asupra concordanțelor) și cînd  $\text{Cov}(X, Y) > 0$ , corelația se zice pozitivă.

Prin urmare, noțiunea de corelație a variabilelor statistice se poate introduce cu ajutorul mărimii covarianță, a cărei semnificație am sugerat-o în rîndurile de mai sus. Deși covarianța indică existența sau inexistența corelației, precum și sensul acesteia (cînd există), ea nu poate juca totuși rolul unui coeficient de corelație pentru motivul că depinde de unitățile de măsură ale celor două variabile și nu are limite fixe de variație, deci nu îndeplinește două dintre condițiile fundamentale ale oricărui coeficient.

Există o procedură foarte simplă de a înlătura neajunsurile menționate. Este suficient să se împartă covarianța la abaterile standard ale lui  $X$  și  $Y$  și se va obține coeficientul dorit. Prin urmare, formula coeficientului de corelație, notat cu  $r$ , este:



$$r = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y}$$

sau

$$r = \frac{\sum_{i=1}^n \sum_{j=1}^t k_{ij}(x_i - \bar{x})(y_j - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \quad (1)$$

Se vede că prin împărțirea cu  $\sigma_x \cdot \sigma_y$  dispăre dimensiunea mărimii și se poate demonstra că:

$$-1 \leq r \leq +1$$

Pentru calcularea coeficientului de corelație cu ajutorul formulei (1), trebuie efectuate succesiv următoarele operații:

1. Se calculează valorile medii,  $\bar{x}$  și  $\bar{y}$ , ale celor două variabile.
2. Se calculează abaterile standard  $\sigma_x$  și  $\sigma_y$ .
3. Se scade  $\bar{x}$  din fiecare valoare  $x_i$  și  $\bar{y}$  din fiecare  $y_j$ .
4. Se efectuează produsele  $k_{ij}(x_i - \bar{x})(y_j - \bar{y})$ , adică, urmărind frecvențele interioare ale tabelului,  $k_{ij}$ , acestea se înmulțesc cu diferențele obținute la punctul 3, diferențe ce caracterizează fiecare linie și coloană a tabelului.
5. Se efectuează suma algebrică (cu semn) a mărimilor de la punctul 4, obținându-se astfel numărătorul coeficientului de corelație.
6. Se calculează numitorul:  $n \sigma_x \sigma_y$ .
7. Se efectuează împărțirea.

Se observă că cele mai mari dificultăți apar la punctele 4 și 5, deoarece, de regulă, valorile medii calculate sînt sub formă de numere cu zecimale și, prin urmare, și diferențele vor fi la fel, ceea ce conduce la un volum mare de muncă pentru efectuarea înmulțirilor. De exemplu, dacă valorile lui  $X$  sînt 1,2,3,4 și 5 iar media iese 2,75, atunci diferențele vor fi:



-1,75; -0,75; +0,25; +1,25 și 2,25. Pentru Y se vor întâmpla foarte probabil lucruri similare și atunci înmulțirea acestor numere cu zecimale nu este o treabă foarte plăcută chiar dacă folosim un minicalculator de birou.

Pentru ușurarea calculelor se pot folosi și alte formule, echivalente cu (1). Dacă efectuăm operațiile de la numărător în formula (1) - deschidem parantezele - vom ajunge la următoarea formă a coeficientului de corelație, reducând termenii asemenea:

$$r = \frac{\sum_{i=1}^s \sum_{j=1}^t k_{ij} x_i y_j - n \bar{x} \bar{y}}{n \cdot \sigma_x \cdot \sigma_y} \quad (2)$$

Folosirea formulei (2) - în care se vede că frecvențele  $k_{ij}$  trebuie înmulțite direct cu valorile corespunzătoare  $x_i$  și  $y_j$  iar mediile apar doar într-o singură operație de înmulțire - se recomandă atunci când valorile  $x_i$  și  $y_j$  sînt mici, deci cînd produsele  $k_{ij} x_i y_j$  nu ne duc la cifre astronomice.

O altă modalitate de simplificare a calculelor este cea care face apel la utilizarea unor valori auxiliare,  $a$  și  $b$  în locul valorilor medii  $\bar{x}$  și  $\bar{y}$ . Procedura este identică cu cea expusă cu ocazia prezentării abaterii standard și, de fapt, aceste valori vor fi folosite și aici în calculul celor două abateri standard dar și la suma dublă de la numărător. Formula poate fi scrisă astfel:

$$r = \frac{\sum_{i=1}^s \sum_{j=1}^t k_{ij} (x_i - a)(y_j - b) + n d_x d_y}{n \cdot \sigma_x \sigma_y} \quad (3)$$

unde

$$d_x = \bar{x} - a$$

$$d_y = \bar{y} - b$$

Valorile  $a$  și  $b$  se aleg și acum de așa manieră încît să se obțină cifre cît mai mici pentru diferențele  $x_i - a$  și  $y_j - b$ . Dacă  $a$  este chiar una dintre valorile lui X și  $b$  una dintre cele ale lui Y, atunci, în plus, pentru o linie și o coloană produsele vor fi nule.



Voi ilustra cu un exemplu mersul calculelor atunci cînd se utilizează formula (3). Să presupunem că într-o cercetare pe un lot de 1000 de muncitori am introdus două variabile: "vîrsta" (X) și "retribuția" (Y). Ambele variabile sînt cu intervale de grupare, fiecare interval fiind apoi caracterizat de o valoare numerică. Distribuția indivizilor după cele două variabile este dată în tabelul 2, în care sînt trecuți și valorile caracteristice intervalelor de grupare.

X \ Y						Total
	2500 lei	2750 lei	3000 lei	3250 lei	3500 lei	
20 ani	50	30	15	5	-	100
30 ani	25	70	30	15	10	150
40 ani	15	50	200	60	25	350
50 ani	10	30	40	100	20	200
60 ani	-	20	15	70	95	200
Total	100	200	300	250	150	1000

**Tabelul 2. Corelația între vîrsta (X) și retribuiția (Y) ale unui lot de 1000 muncitori (date fictive)**

Se observă că, pe măsură ce avansăm pe scara vîrstei, ponderea indivizilor cu trepte mai înalte de retribuiție crește. Prin urmare, tabelul 2 redă un caz tipic de corelație pozitivă iar calculul coeficientului va arăta intensitatea acestei corelații.

Alegem următoarele valori ajutătoare:

$$a = 40 \text{ ani}$$

$$b = 3000 \text{ lei}$$

și aranjăm datele pentru calcularea mediilor și abaterilor standard sub forma cunoscută:



Pentru  $X_i$

$x_i$	$x_i - a$	$k_i$	$k_i(x_i - a)$	$(x_i - a)^2$	$k_i(x_i - a)^2$
20	-20	100	-2.000	400	40.000
30	-10	150	-1.500	100	15.000
40	0	350	-	-	-
50	10	200	2.000	100	20.000
60	20	200	4.000	400	80.000
Total		1000	2.500	1000	155.000

Rezultă imediat:

$$\begin{aligned} d_x &= 2,5 & d_x^2 &= 6,25 & \bar{x} &= 42,5 \\ s_x^2 &= 155 & \sigma_x^2 &= 148,75 & \sigma_x &= 12,2 \end{aligned}$$

Pentru  $Y_i$

$y_j$	$y_j - b$	$k_{.j}$	$k_{.j}(y_j - b)$	$(y_j - b)^2$	$k_{.j}(y_j - b)^2$
2.500	-500	100	-50.000	250.000	25.000.000
2.750	-250	200	-50.000	62.500	12.500.000
3.000	0	300	-	-	-
3.250	250	250	62.500	62.500	15.625.000
3.500	500	150	75.000	250.000	37.500.000
Total		1000	37.500	625.000	90.625.000

Si deci:

$$\begin{aligned} d_y &= 37,5 & d_y^2 &= 1406,25 & \bar{y} &= 3.037,5 \\ s_y^2 &= 90.625 & \sigma_y^2 &= 89.218,75 & \sigma_y &= 298,7 \end{aligned}$$

Trecem apoi la calcularea sumei de la numărător. In schema de calcul ce urmează am așezat pe prima coloană frecvențele interioare ale tabelului de corelație, urmărindu-le pe linii.



Pentru produsele finale am construit două coloane, pentru a trece separat valorile pozitive și negative.

$k_{ij}$	$x_i - a$	$y_j - b$	$k_{ij}(x_i - a)(y_j - b)$
			+
			-
50	-20	-500	500.000
30	-20	-250	150.000
15	-20	0	0
5	-20	250	25.000
0	-20	500	0
-----			
25	-10	-500	125.000
70	-10	-250	175.000
30	-10	0	0
15	-10	250	37.500
10	-10	500	50.000
-----			
15	0	-500	0
50	0	-250	0
200	0	0	0
60	0	250	0
25	0	500	0
-----			
10	10	-500	50.000
30	10	-250	75.000
440	10	0	0
100	10	250	250.000
20	10	500	100.000
-----			
0	20	-500	0
20	20	-250	100.000
15	20	0	0
70	20	250	350.000
95	20	500	950.000

T O T A L : 2.600.000 -337.000

Deci valoarea sumei duble de la numărător va fi:



$$= 264 =$$

$$2.600.000 - 337.500 = 2.262.500$$

Urmează apoi că:

$$n \cdot d_x d_y = 93.750$$

și deci valoarea numărătorului va fi:

$$2.262.500 - 93.750 = 2.168.750$$

Calcululele de la numitor ne conduc la:

$$n \cdot \sigma_y \cdot \sigma_y = 1000 \cdot 12,2.298,7 = 3.644.140$$

și deci

$$r = \frac{2.168.750}{3.644.140} = 0,60$$

Să prezentăm în continuare câteva proprietăți ale coeficientului de corelație, însoțite și de unele precisări.

a) Așa cum am spus, intervalul de variație pentru  $r$  este de la  $-1$  la  $+1$ . În cazul în care se ating aceste valori extreme, corelația poate fi numită perfectă (pozitivă sau negativă). Valoarea nulă a coeficientului indică o situație în care cele două variabile sînt necorelate.

b) O condiție suficientă pentru ca  $r$  să atingă valoarea  $1$  este ca cele două variabile să ia aceleași valori, ca distribuțiile lor de frecvență să coincidă (adică frecvențele marginale ale tabelului de corelație să fie identice) și ca toate frecvențele interioare ale tabelului să se afle pe diagonala principală. Sau, schematic:

$$\left. \begin{array}{l} s = t \\ x_1 = y_1 \\ k_{.1} = k_{1.} \\ k_{11} = k_{.1} \end{array} \right\} \Rightarrow r = 1$$

Proprietatea reciprocă nu este întotdeauna adevărată, în sensul că există situații în care valorile  $x_1$  și  $y_1$  nu coincid,



dar coeficientul ia valoarea 1. In orice caz, corelația perfectă nu se poate atinge decât dacă tabelele sînt pătratice ( $s=t$ ) și toate frecvențele interioare sînt diagonale.

a) Dacă frecvențele interioare,  $k_{ij}$ , verifică relația de independență:

$$k_{ij} = \frac{k_{i.} \cdot k_{.j}}{n}$$

atunci valoarea lui  $r$  este nulă. Reciproca nu este nici aici adevărată, putîndu-se obține valori nule pentru  $r$  și cînd frecvențele nu verifică relația de independență. Prin urmare, independența implică neocorelația variabilelor sau, invers, existența corelației implică asocierea (în sensul definit la caracteristicile calitative). Din faptul că afirmațiile reciproce nu sînt adevărate, rezultă cele ce spuneam la început, și anume că cele două noțiuni aplicabile în cazul variabilelor (asocierea și corelația) nu se suprapun exact.

d) O altă proprietate remarcabilă a coeficientului de corelație constă în faptul că valoarea lui nu este, la drept vorbind, o funcție de valorile variabilelor, ci de diferențele dintre aceste valori, luate două cîte două, deci de distanțele dintre indivizi și mai exact de rapoartele între aceste distanțe. Asta înseamnă că putem deplasa spre stînga sau spre dreapta seria de valori și, de asemenea, că putem înmulți sau împărți cu un același număr valorile unei variabile fără a se modifica valoarea lui  $r$ . Sau, altfel spus, coeficientul de corelație rămîne constant dacă efectuăm transformări lineare ale celor două variabile:

$$X' = a + bX$$

$$Y' = c + dY$$

Importanța proprietății de față nu constă numai în faptul că prin asemenea transformări putem ușura în mod simțitor calculele, destul de laborioase altfel, ce intervin în formule (1), (2) și (3). Ea devine evidentă pentru situațiile în care procedura de atribuire a valorilor nu este cu totul riguroasă, acesta fiind un caracter aproape constant al variabilelor



sociale. Stim astfel că, pentru a ajunge la un "adevărat" coeficient de corelație între două variabile, nu este necesar să măsurăm exact valorile lor; va fi suficient să găsim o procedură ce permite atribuirea valorilor și care să reproducă în mod corect raportul intervalelor. Dacă notăm cu  $1, 2, \dots, s$  valorile caracteristicii, important, pentru analiza de corelație, nu este faptul că indivizii iau sau nu aceste valori și nici că diferența dintre valoarea unei clase și cea a precedentei este unitatea, ci numai că distanțele între indivizii aparținând la două clase consecutive sînt constante.

Revenind la problematica simplificării calculelor, menționez un caz destul de frecvent întîlnit, acela în care valorile uneia sau ambelor variabile sînt în progresie aritmetică. Atunci aceste valori se pot transforma în mulțimea numerelor naturale sau întregi, păstrînd caracterul de progresie aritmetică. De exemplu, în exercițiul prezentat mai sus, atît pentru valorile vîrstei cît și pentru cele ale retribuției putem folosi numere întregi în valoare absolută mult mai mică. Cea mai importantă simplificare se obține atunci cînd o valoare mijlocie se transformă în zero iar celelalte în  $\pm 1, \pm 2$ . În exemplul nostru putem lua ca valori pentru ambele variabile următoarele:  $+2, -1, 0, 1$ , și  $2$ . urmînd apoi să aplicăm direct formula (2) în calculul lui  $r$ . Să reluăm acest exemplu:

Pentru X :

$x_i$	$k_i$	$k_i \cdot x_i$	$x_i^2$	$k_i \cdot x_i^2$
-2	100	-200	4	400
-1	150	-150	1	150
0	350	-	-	-
1	200	200	1	200
2	200	400	4	800
<hr/>				
Total	1000	250		1550



Luind pe 0 valoare ajutătoare, rezultă:

$$\begin{aligned} \bar{x} &= 0,25 & d_x &= 0,25 & d_x^2 &= 0,0625 \\ s_x^2 &= 1,55 & \sigma_x^2 &= 1,4875 & \sigma_x &= 1,22 \end{aligned}$$

Pentru Y:

$y_j$	$k_{.j}$	$k_{.j}y_j$	$y_j^2$	$k_{.j}y_j^2$
-2	100	-200	4	400
1	200	200	1	200
0	300	-	-	-
1	250	250	1	250
2	150	300	4	600
<hr/>				
Total	1000	150		1450

81

$$\begin{aligned} \bar{y} &= 0,15 & d_y &= 0,15 & d_y^2 &= 0,0225 \\ s_y^2 &= 1,45 & \sigma_y^2 &= 1,4275 & \sigma_y &= 1,19 \end{aligned}$$

Calculul sumei duble este prezentat mai jos, fără a mai reproduce cazurile în care produsele sînt nule.

$k_{ij}$	$x_i$	$y_j$	$k_{ij}x_iy_j$	-
50	-2	-2	200	
30	-2	-1	60	
5	-2	1	10	
25	-1	-2	50	
70	-1	-1	70	
15	-1	1	15	
10	-1	2	20	
10	1	-2	20	
30	1	-1	30	
100	1	1	100	
20	1	2	40	



20	2	-1	40
70	2	1	140
95	2	2	380
<hr/>			
TOTAL:		1040	-135 = 905

Deci suma dublă este 905 și urmează să calculăm:

$$n \cdot \bar{x} \cdot \bar{y} = 1000 \cdot 0,25 \cdot 0,15 = 37,5$$

ceea ce înseamnă că numărătorul coeficientului va fi:

$$905 - 37,5 = 867,5$$

Pentru numitor, înmulțim

$$n \cdot \sigma_x \cdot \sigma_y = 1000 \cdot 1,22 \cdot 1,19 = 1451,8$$

și deci

$$r = \frac{867,5}{1451,8} = 0,60$$

adică s-a ajuns la același rezultat cu un volum mai mic de calcule.

Desigur, că în condițiile în care cercetătorul dispune de un calculator electronic pentru prelucrarea datelor, totuși această poveste despre simplificarea operațiilor nu-și mai are rostul. Am ținut totuși să ofer procedeele respective pentru motivul că încă multe calcule se mai efectuează manual. Procedăm așa mai ales atunci când calcularea unui coeficient de corelație (sau a unui alt indicator) este o problemă în sine și imediată, deci când nu este inclusă într-un program mai amplu de prelucrare a datelor, pentru care utilizarea calculatorului să fie dacă nu indispensabilă cel puțin foarte eficientă. Anumite dificultăți legate de accesul la mașinile electronice de calcul, lipsa de pregătire a majorității oamenilor pentru a purta un dialog eficient cu computerul și altele fac însă rentabile unele calcule cu mașini de birou, operații ce nu depășesc un volum de muncă de câteva ore. Or, aflarea unui sau a câtorva coeficienți de corelație, chiar pentru variabile mai complexe decât cele din exemplul dat, este o muncă ce se încadrează în acest quantum de timp.



Un caz particular. Dacă problema asocierii se pune doar atunci când avem observații asupra unui număr mare de indivizi statistici - numai în situații cu totul deosebite, un tabel de asocieri, chiar și pentru caracteristici dihotomice, spune ceva dacă cuprinde mai puțin de 100 de cazuri - corelația poate fi urmărită iar coeficientul de corelație calculat, și pentru loturi mici de indivizi. În asemenea situații, nu mai are rost construcția caracteristicilor cu intervale de grupare, deoarece este mai comodă însușirea indivizilor într-o anumită ordine și evidențierea valorilor pe care le ia fiecare în raport cu X și Y. Atunci, în formulele care dau coeficientului, vor dispărea frecvențele  $k_{ij}$ , însumarea se va face urmărindu-se indivizii așezați într-o anumită ordine și, prin urmare, suma dublă se va transforma în una simplă, după un singur indice, mergând de la 1 la  $n$ . Astfel, formula (1) va deveni:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n \cdot \sigma_x \sigma_y} \quad (4)$$

Și în mod similar se pot scrie și celelalte două.

Să luăm pentru exemplificare o situație foarte simplă. Să presupunem că pentru 10 județe - pe care le-am simbolizat cu litere de la A la J - dorim să vedem ce corelație există între numărul cinematografelelor care funcționează în fiecare județ (X) și numărul spectatorilor de film într-un an (Y). Cifrele sînt și de această dată fictive deoarece pentru scopuri didactice sînt preferabile exemple special construite.

	A	B	C	D	E	F	G	H	I	J	Total
Județul											
Nr. de cinemato- grafe(X)	100	120	130	140	150	160	170	190	210	230	1600
Nr. spec- tatori, în mili- oane(Y)	3,1	2,4	6,4	2,8	5,9	4,0	5,6	5,0	6,8	8,0	50

Tabel 3. Numărul cinematografelelor și spectatorilor de film în 10 județe (cifre fictive)



Calcululele se vor efectua în felul următor:

Jud.	$x_i$	$y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
A	100	3,1	-60	-1,9	3600	3,61	114
B	120	2,4	-40	-2,6	1600	6,76	104
C	130	6,4	-30	1,4	900	1,96	42
D	140	2,8	-20	-2,2	400	4,84	44
E	150	5,9	-10	0,9	100	0,81	9
F	160	4,0	0	-1,0	0	1,00	0
G	170	5,6	10	0,6	100	0,36	6
H	190	5,0	30	0	900	0	0
I	210	6,8	50	1,8	2500	3,24	90
J	230	8,0	70	3,0	4900	9,00	210
Tot.	1600	50,0			15000	31,58	568 - 51 = 517

Din primele două sume, rezultă evident că:

$$\bar{x} = 160; \quad \bar{y} = 5$$

valori cu ajutorul cărora s-au calculat expresiile din coloanele următoare. (Dacă n-ar fi ieșit cifre rotunde, se puteau folosi valori ajutoare ca și mai înainte). Urmează apoi că:

$$\sqrt{x} = \sqrt{1500} = 38,73$$

$$\sqrt{y} = \sqrt{3,158} = 1,78$$

În fine, ultima coloană dă numărătorul coeficientului de corelație și:

$$r = \frac{517}{10 \cdot 38,73 \cdot 1,78} = 0,75$$

Am obținut deci o valoare pozitivă ridicată, ce ilustrează o legătură strinsă între numărul cinematografelor și cel al spectatorilor, în județele analizate. Urmărind ultima coloană, observăm că doar două din cele zece produse sînt negative (pentru județele C și E). avînd de a face aici cu ceea ce numeam la început discrepanță de comportament, în ambele cazuri numărul de



cinematografe fiind sub media lotului analizat iar numărul spectatorilor peste medie.

Doresc să atrag atenția că foarte des în texte - așa cum am făcut și eu mai sus - mărimea coeficienților de corelație este apreciată calitativ: "valoare ridicată", "coeficient foarte mare", "corelație slabă" etc. Menționez că nu există, niste limite stabilite între care să se găsească valoarea coeficientului pentru a i se atribui asemenea etichete calitative. Aprecierile depind de cele mai multe ori de așteptările cercetătorului, iar nu o dată se întâmplă ca un același coeficient să fie apreciat de unii drept mare și de alții mic. Chiar și în exemplul anterior, având de a face cu două variabile care în mod "normal" trebuie să fie corelate, valoarea 0,75 poate fi contradictoriu apreciată, după cum este sub sau peste așteptările celui care efectuează analiza.

În sfârșit, mai menționez că adesea se vorbește despre valori "semnificative" sau "nesemnificative" ale lui  $r$ . Se cuvine subliniat că termenul de "semnificativ" are în statistică o semnificație foarte clară, referindu-se exclusiv la mărimile obținute asupra eșantioanele probabiliste. De pildă, în exemplul precedent, dacă lo județe constituie obiectul exclusiv al studiului nostru, coeficientul obținut (chiar dacă ar fi fost numai 0,05 să zicem) nu este nici semnificativ nici nesemnificativ; el este atît cît este. Dacă însă, cele lo județe ar fi fost considerate un eșantion din toate județele țării, atunci problema semnificației lui  $r$  s-ar pune, avînd menirea să dea răspuns la întrebarea dacă nu cumva, prin selecție, s-a obținut din întâmplare o valoare diferită de zero. Există proceduri statistice bine elaborate pentru a soluționa o asemenea chestiune.

## 2. Corelația rangurilor

Problema corelației rangurilor este, în aparență, diferită prin conținut de cele discutate aici, locul său pîrînd mai degrabă a fi în capitolul ce tratează asocierea. În esență, e vorba despre următorul lucru. Să presupunem că avem un grup relativ mic de indivizi și două criterii de ordonare (clasificare ierarhică) a acestora. Se pune problema de a găsi un coeficient care să exprime gradul de similitudine al ordonării după cele două criterii.



Putem să transcriem aceasta într-un limbaj mai apropiat de cel utilizat în cursul de față. Ordonarea indivizilor înseamnă, de fapt, aplicarea unei scale ordinale pentru o caracteristică dată. În cazul unei ordonări stricte, fiecare treaptă a scalei va conține numai un singur individ din populație, deci scala are atâtea trepte câți indivizi sînt. Să presupunem, momentan, că ne găsim într-o asemenea situație simplă. Atunci apare cu claritate că problema noastră este aceea a calculării unui coeficient de corelație între două caracteristici calitative, fapt care intră în contradicție cu precizările anterioare, și anume că vom vorbi de corelație doar în cazul caracteristicilor cantitative. Vom vedea imediat cum se înlătură acest paradox.

Să începem prin a clarifica mai mult datele problemei. Fie o mulțime de  $n$  indivizi statistici pe care să-i botezăm cu:

$$A_1, A_2, \dots, A_n$$

unde indicii  $1, 2, \dots, n$  nu sînt decît niște etichete pentru identificare și pentru a arăta că în total sînt  $n$  indivizi. Ordinea scrierii lor nu are, așadar, nici o importanță. Să presupunem că, în raport cu un criteriu  $X$ , am reușit să realizăm o ordonare strictă a indivizilor, de la cel care posedă proprietatea în cea mai mare măsură la cel care o posedă în cea mai mică, de pildă. Aceasta nu înseamnă altceva decît a stabili o corespondență biunivocă între mulțimea indivizilor și o mulțime  $\{x_1, x_2, \dots, x_n\}$  formată din numere naturale diferite, cuprinse între 1 și  $n$  (o permutare a numerelor  $1, 2, \dots, n$ ). Astfel, dacă lui  $A_1$  îi vom pune în corespondență numărul  $x_1$  egal cu 3 înseamnă că individul  $A_1$  se află pe locul 3 al clasamentului stabilit după  $X$  ș.a.m.d. Procedînd similar cu  $Y$ , obținem un alt clasament, respectiv o altă permutare a aceleiași mulțimi de numere:  $\{y_1, y_2, \dots, y_n\}$ . Finalmente vom ajunge la un tabel de forma:

Indivizi	$A_1$	$A_2$	...	$A_n$
Rang după $X$	$x_1$	$x_2$	...	$x_n$
Rang după $Y$	$y_1$	$y_2$	...	$y_n$



Pentru a înțelege perfect despre ce este vorba, că dăm un exemplu foarte la îndemână. Fie mulțimea indivizilor formată din cele 18 echipe de fotbal care participă într-un an la campionatul primei divizii și fie  $x_1, x_2, \dots, x_{18}$  locul pe care fiecare echipă îl ocupă în clasament la sfârșitul turului iar  $y_1, y_2, \dots, y_{18}$  locurile ocupate de același echipe la terminarea campionatului. Un coeficient de corelație a rangurilor are, cum spunem, rolul să indice similitudinea, apropierea dintre cele două clasamente. O similitudine perfectă - adică, în cazul nostru, menținerea de către fiecare echipă a locului în clasament - va trebui să ne conducă la o valoare +1 a coeficientului, iar o discrepantă totală - adică inversarea clasamentului în varianta a doua - , la o valoare -1. Mărimea nulă va trebui să se atingă atunci când între cele două clasamente nu există nici o legătură, respectiv când unul dintre ele dispune aleator indivizii în raport cu celălalt.

Mărimea care satisface aceste cerințe se notează cu  $\rho$  și se numește coeficientul de corelație a rangurilor sau coeficientul lui Spearman. Formula sa este:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

unde cu  $d_i$  am notat diferențele rangurilor (pozițiilor) fiecărui individ în cele două clasamente:

$$d_i = x_i - y_i ; i = 1, 2, \dots, n$$

Practic, calculul lui  $\rho$  este foarte simplu. Datele se trec într-un tabel ca cel de mai sus, doar că se scriu pe verticală (pentru a se facilita însumarea) și se mai adaugă două coloane: una pentru valorile  $d_i$ , alta pentru  $d_i^2$ .

Să reluăm exemplul cu județele și în loc să folosim efectiv numărul cinematografelelor și cel al spectatorilor să ierarhizăm județele după aceste două criterii (de fapt, se vede că în raport cu prima variabilă județele au și fost scrise în ordine) Vom ajunge la situația din tabelul 4.



Jude- țul	Rang după X	Rang după Y	$d_1$	$d_1^2$
A	10	8	2	4
B	9	10	-1	1
C	8	3	5	25
D	7	9	-2	4
E	6	4	2	4
F	5	7	-2	4
G	4	5	-1	1
H	3	6	-3	9
I	2	2	0	0
J	1	1	0	0
TOTAL				52

**Tabelul 4. Ierarhia județelor după variabilele  
din tabelul 3.**

Deci suma pătratelor diferențelor este 52, și aplicînd formula:

$$\xi = 1 - \frac{6 \cdot 52}{1000 - 10} = 1 - \frac{312}{990} = 1 - 0,32$$

$$\xi = 0,68$$

După efectuarea acestor calcule intuim mai ușor că  $\xi$  satisface proprietățile cerute. Dacă cele două clasamente sînt identice, atunci toate diferențele vor fi nule și deci tot nul va fi și termenul care se scade din 1. Cu cît cele două clasamente sînt mai deosebite unul de celălalt, cu atît diferențele vor fi mai mari și deci termenul care se scade mai mare. La atingerea de către acesta a valorii 1 ( $\xi = 0$ ), clasamentele sînt independente, iar la atingerea lui 2 ( $\xi = -1$ ) clasamentele se inversează.



Să mai luăm un exemplu. Să presupunem că doi profesori (de pildă, unul de muzică și altul de educație fizică) au fost solicitați să ordoneze 15 copii în funcție de aptitudinile la disciplinele respective. Clasamentele întocmite de către cei doi profesori (imaginari) apar în tabelul 5, ca și restul elementelor pentru calcularea lui  $S$ .

Din datele tabelului 5, se obține:

$$S = 1 - \frac{6.560}{3360} = 1 - \frac{3360}{3360} = 1 - 1 = 0$$

Deci se poate afirma că, în lotul studiat, aptitudinile la cele două obiecte sînt aleator distribuite, adică nu se manifestă nici o tendință ca cei care sînt mai buni la muzică să fie mai buni ori mai slabi la sport.

Elevii	Clasament la muzică (X)	Clasament la sport (Y)	$d_1$	$d_1^2$
A	7	10	-3	9
B	3	11	-8	64
C	1	5	-4	16
D	4	8	-4	16
E	10	2	8	64
F	15	6	8	64
G	12	15	-3	9
H	11	1	10	100
I	8	3	5	25
J	2	9	-7	49
K	6	4	2	4
L	5	12	-7	49
M	9	14	-5	25
N	13	13	0	0
O	14	7	7	49
TOTAL:			560	

Tabelul 5. Clasamentele celor 15 elevi, după aptitudinile la două discipline



Coeficientul de corelație a rangurilor este foarte des utilizat în științele umane (mai ales în psihologie, unde prin tradiție se lucrează cu egantioane mici) și nu o dată cei care scriu despre el îi laudă virtuțile. Fascinația pe care o exercită acest coeficient se datorează iluziei că el ar reprezenta o măsură a unei corelații între caracteristici calitative, care deci nu implică operația de măsurare. Se poate arăta ușor că această iluzie se spulberă imediat dacă ne dăm osteneala să înțelegem de unde provine formula lui Spearman.

Nu este locul să prezint aici deducția matematică a acestei formule. Este suficient să precizez că, de fapt, coeficientul  $S$  nu este decât un obșnuit coeficient de corelație  $r$ , obținut prin aplicarea formulei (4) a lui  $r$ , în care valorile  $x_i$  și  $y_i$  sînt numere naturale distincte între 1 și  $n$ . Datorită faptului că în toate situațiile valorile lui  $X$  și  $Y$  sînt mereu cele amintite mai sus și că fiecare valoare este luată o singură dată este clar că se va ajunge la o formulă de simplificare a coeficientului. De pildă, mediile lui  $X$  și  $Y$  vor fi egale și constante, pentru  $n$  dat, indiferent de clasamente, căci:

$$\bar{x} = \bar{y} = \frac{1+2+\dots+n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}$$

La fel vor fi egale și constante și abaterile standard iar suma de la numărătorul coeficientului se va exprima în funcție de  $n$  și de diferențele  $d_i$ .

Esențial este deci faptul că  $S$  este tot un coeficient de corelație, făcînd în consecință, apel tot la valori ale variabilelor. Scamatoria ni se dezvăluie atunci imediat: s-a utilizat pentru fiecare variabilă o clasificare; indivizii au primit un număr care le desemna locul în clasament iar apoi acest număr a fost luat drept valoare. S-a efectuat deci o trecere de la numere ordinale la numere cardinale.

Gravitatea "infracțiunii" este substanțial diminuată de proprietatea coeficientului de corelație menționată la punctul



d), în paragraful precedent. Într-adevăr, atribuirea valorilor  $1, 2, \dots, n$  nu înseamnă - în calculul lui  $\rho$  - că indivizii iau efectiv aceste valori, ci numai că distanțele dintre ei sînt egale. Drept urmare, coeficientul de corelație a rangurilor este efectiv un coeficient de corelație, aplicat atunci cînd nu este posibilă operația de măsurare dar se pot ierarhiza indivizii studiați, presupunînd că diferențele de despart doi indivizi consecutivi sînt identice.

Se înțelege atunci că, ori de cîte ori este posibil, trebuie preferat calculul lui  $r$  și nu cel al lui  $\rho$  (lucrul fiind evident, de pildă, în exemplul cu județele). Cerința unor metodologi de a se folosi corelația rangurilor atunci cînd măsurătorile sînt imprecise nu este totdeauna justificată. Să presupunem că dorim să corelăm două variabile cum ar fi pregătirea unor elevi la două obiecte de învățămînt. Elevii avînd note la respectivele materii corelația se poate calcula folosind notele în calitate de valori sau cu ajutorul coeficientului lui Spearman, după clasamentele ce se pot întocmi pe baza acelorasi note. Unii pedagogici preferă aceasta a doua cale pentru motivul că notele nu reprezintă o măsură exactă a nivelului de cunoștințe al elevilor. Ei nu înțeleg că aceeași imprecizie poate duce la o clasificare greșită. Mai mult, să zicem că la un obiect elevii L, M și N vor avea notele 8, 7,50 și respectiv, 7,45 apărînd în această ordine în clasament. Cele 5 sutimi de punct ce-l despart pe N de M pot fi datorate unei întîmplări și este foarte greu să acceptăm că distanța dintre L și M va fi egală cu cea dintre M și N. Or, în calculul coeficienților de corelație distanțele contează în primul rînd.

Ranguri intrunite. Se știe că uneori clasamentele nu se pot întocmi în mod perfect, adică 2,3 sau chiar mai mulți indivizi se pot clasa pe același loc. Atunci, strict vorbind, formula scrisă pentru  $\rho$  nu se mai poate aplica, existînd elaborată o altă care ține seama de numărul rangurilor intrunite sau legate. Cum ea este destul de complicată, nu o voi mai prezenta aici, precizînd doar că, dacă numărul rangurilor intrunite este mic, se poate folosi și formula lui Spearman, acceptînd că aceasta ne dă o aproximație a coeficientului de corelație adevărat.



În astfel de cazuri, trebuie să acordăm atenție felului cum se notează rangurile. Să presupunem că, într-un clasament, după individul care ocupă primul loc urmează doi la egalitate. Ei vor corespunde pozițiilor 2 și 3, dar cum nu-i putem diferenția trebuie să le acordăm aceeași valoare, care va fi, în cazul imaginat, 2.5 iar individului care urmează după cei doi îi vom acorda rangul 4. Deci dacă mai mulți indivizi ocupă aceeași poziție în clasament, fiecare va primi o valoare egală cu rangul mediu, pe când individul care urmează primește aceeași valoare ce i-ar fi revenit dacă precedenții ar fi fost diferențiați.

### 3. Regresie, corelație parțială și multiplă

Si în cazul variabilelor statistice există proceduri și încă mult mai bine puse la punct decât pentru caracteristicile calitative - de analiză a raporturilor simultane dintre mai mulți factori. Este evident o problemă de generalizare a coeficientului de corelație la situația mai complexă când numărul variabilelor studiate depășește pe 2. Pentru a putea înțelege semnificația coeficienților generalizați, chiar și în maniera foarte sumară în care îi vom prezenta în paragraful de față, va trebui să specificăm și alte interpretări ale coeficientului simplu de corelație, interpretări ce se desprind din analiza noțiunii de regresie. Ar fi greșit însă să se înțeleagă că regresia este doar o procedură statistică ce ne ajută să facem trecerea amintită; dimpotrivă, analiza de regresie reprezintă un capitol fundamental al statisticii parametrice, important în sine, important și în elucidarea raportului dintre două variabile, lucru care cred că se va înțelege din rândurile ce urmează.

Regresia în cazul a două variabile. Fie  $X$  și  $Y$  două variabile statistice, definite pe o mulțime de  $n$  indivizi. Pentru a facilita înțelegerea rândurilor ce urmează, să presupunem că acest număr este mic, deci nu efectuăm restrângeri pe intervale de grupare și evidențiem fiecare individ prin valorile  $x_i$  și  $y_i$  pe care le ia, supoziție prin care nu se restrânge deloc generalitatea problemei.

Putem da fenomenului o ilustrare grafică foarte simplă.



Dacă luăm în plan două axe de coordonate ortogonale și marcăm pe una valorile în raport cu X și pe cealaltă în raport cu Y, atunci fiecare individ primește ca imagine un punct din plan de coordonate  $(x_1, y_1)$ . Imaginea de ansamblu a tuturor celor  $n$  indivizi studiați se va prezenta ca un "nor" de asemenea puncte.

Se poate ușor înțelege că forma acestui "nor" ne va da o indicație asupra existenței și sensului unei legături de corelație între cele două variabile, după tendințe de cuplare a valorilor lui X cu cele ale lui Y. Așa de pildă, în cazul unei corelații pozitive, când deci valorile mari ale lui X se leagă, în medie, cu valori mari ale lui Y și cele mici cu cele similare, <sup>norul</sup> de puncte va fi alungit pe o direcție apropiată de cea a bisectoarei întâi a sistemului de axe (cea care trece prin cadranele I și III) iar în caz contrar, se va contura un nor alungit după o direcție similară cu cea a bisectoarei a doua. Cele două cazuri sînt redată în figurile 1 și 2.

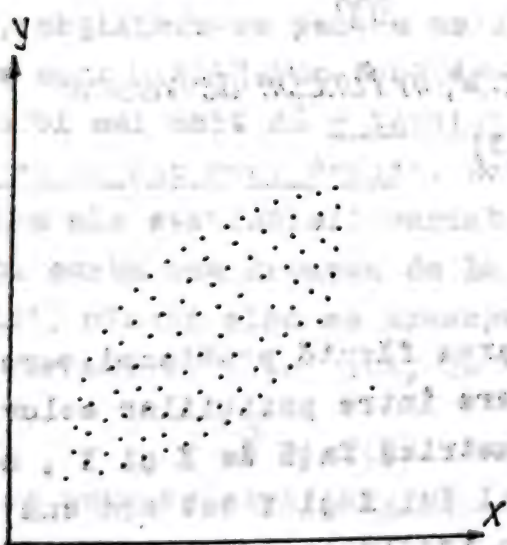


Fig.1 Nor de puncte sugerind o corelație pozitivă

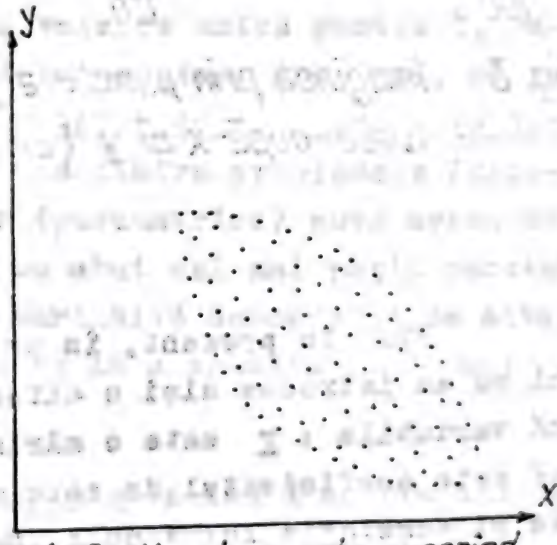


Fig.2 Nor de puncte sugerind o corelație negativă

Ca să prezentăm o situație și mai concretă, să reluăm exemplul cu județele utilizat deja în acest capitol. Reprezentînd grafic situația din tabelul 3, vom ajunge la imaginea din figura 3.



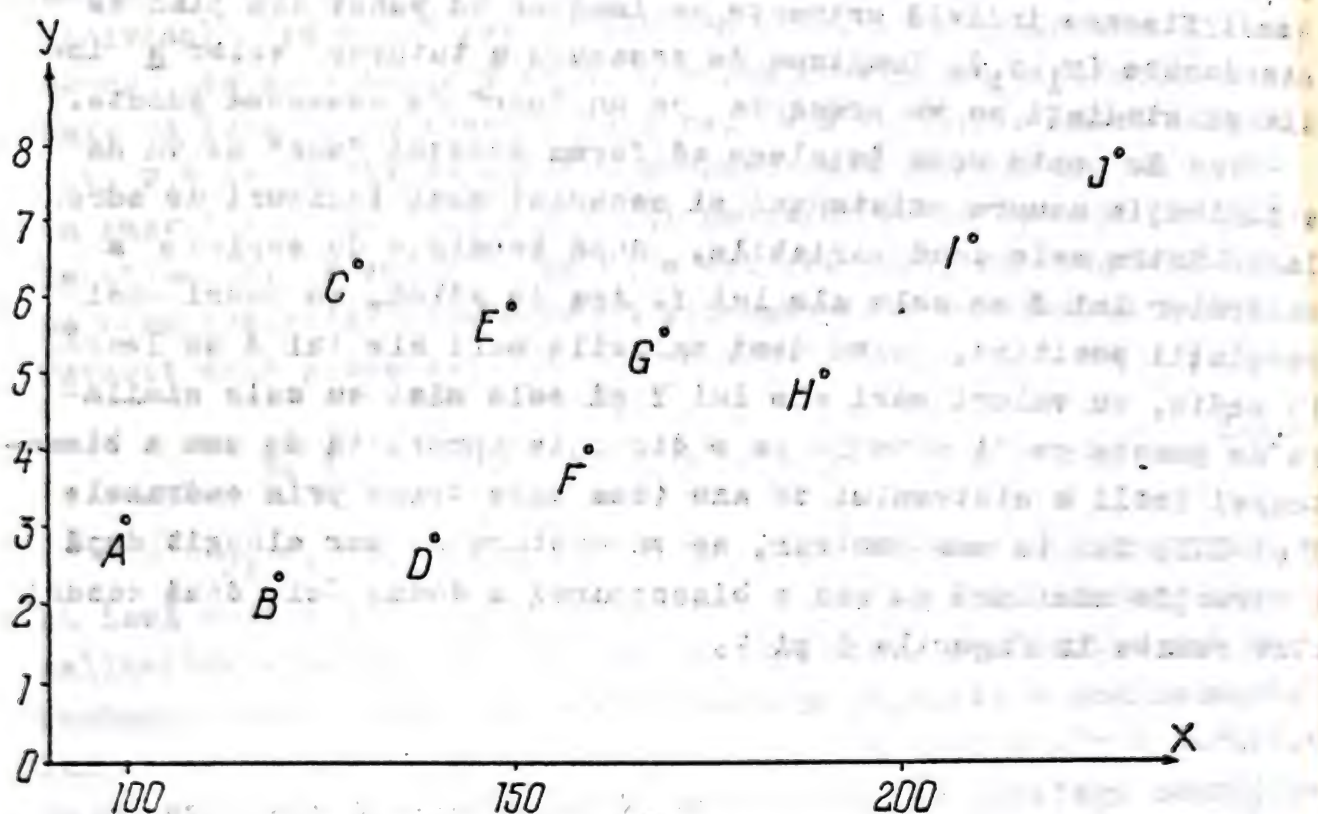


Fig. 3 Imaginea prin puncte a județelor, în funcție de valorile luate după  $x$  și  $y$  (datele tab.3)

Pînă în prezent, în prezentarea făcută problemei corelației nu am introdus nici o diferențiere între pozițiilor celor două variabile :  $r$  este o mărime simetrică față de  $X$  și  $Y$ , adică este coeficientul de corelație al lui  $X$  și  $Y$  tot așa cum este și corelația lui  $Y$  cu  $X$ . De acum încolo lucrurile se vor schimba puțin, căci vom presupune una dintre variabile ca fiind dependentă de cealaltă, în sensul matematic al cuvîntului, adică vom zice că variația uneia antrenează variația celeilalte. Să nu uităm încă că ne aflăm în domeniul statisticii. Dacă dependența ar fi funcțională (deci riguroasă), am zice că unei creșteri a lui  $X$  cu o valoare dată îi corespunde o creștere sau o descreștere a lui  $Y$  cu altă valoare dată. În dependențele de tip statistic, formularea poate îmbrăca învelișul verbal următor: unei creșteri a lui  $X$  cu o valoare dată îi corespunde în medie o creștere



a lui Y cu o altă valoare.

Observație. Aceste dependențe formale, indiferent că sînt funcționale sau statistice, nu reflectă în mod obligatoriu dependențe materiale între fenomenele pe care variabilele le reprezintă. Se poate ca, de pildă, unei creșteri a numărului de cinematografe (X) să-i corespundă - așa cum s-a întîmplat în exemplul nostru - o creștere, în medie, a numărului de spectatori (Y). Relația efectivă poate fi în ambele sensuri: fie că numărul sporit al cinematografelor constituie o condiție favorizantă pentru atragerea unui număr mai mare de spectatori, fie că volumul acestor instituții este determinat de efectivul demografic al județelor, deci de numărul potențial de spectatori. Probabil că aceasta a doua alternativă este mai plauzibilă, dar asta nu ne împiedică să-l calculăm pe "Y în funcție de X".

Intuitiv, dependența statistică dintre variabile este redată de forma norului de puncte; dacă dependența ar fi funcțională, atunci punctele s-ar plasa exact pe o dreaptă sau pe o curbă carecure, obținîndu-se pentru un X dat o valoare unică pentru Y, valoare calculabilă din ecuația curbei respective. Așa însă, se poate vorbi cel mult de o tendință a punctelor de a se grupa în jurul unei curbe sau unei drepte. Astfel, una dintre problemele fundamentale ale statisticii variabilelor (parametrice) este aceea de a găsi curba sau dreapta de la care se abat cel mai puțin punctele-imagini, atunci cînd se presupune o variabilă dependentă de alta. Cu X variabilă dependentă, se va ajunge la o ecuație de forma:

$$\hat{Y} = f(X)$$

care se numește ecuația curbei de regresie a lui Y asupra lui X. În particular, cînd

$$\hat{Y} = a + bX \quad (1)$$

se va obține dreapta de regresie.

În continuare mă voi mărgini la cîteva considerații asupra regresiei lineare, datorită faptului că este cazul cel mai simplu și pentru că, în realitate, este și cel mai frecvent întîlnit.

Așadar, avînd mulțimea de puncte reprezentînd indivizii noștri statistici, problema este de a determina o dreaptă de forma (1) de așa manieră încît ea să se afle cît mai aproape de totali-



tatea acestor puncte. Faptul că este important care dintre cele două variabile este socotită independentă și care dependentă va reieși imediat când se va preciza sensul expresiilor "cea mai mică abatere a punctelor de la dreaptă" sau "cea mai apropiată dreaptă de puncte".

Mai întâi, trebuie să înțelegem că determinarea dreptei de regresie ne pune în posesia unei dependențe funcționale între două cîrmi  $X$  și  $Y$ . Cum însă dependența dintre variabilele noastre inițiale nu este, în general, de această formă, devine limpede că variabila  $\hat{Y}$  din ecuația (1) nu este nici ea variabila  $Y$  inițială, ci una nouă ale cărei valori rezultă efectuînd calcule în membrul drept al relației (1), ceea ce explică deci notația folosită. Relația (1) exprimă tendința medie de covariație a lui  $X$  și  $Y$  și ea nu are drept scop să prezică niște coordonate individuale, ci un comportament mijlociu, de grup. Intr-adevăr, dacă  $x_1$  este valoarea pe care un individ  $i$  o ia în raport cu  $X$ , este foarte probabil ca valoarea  $y_1$  pe care același individ o ia față de  $Y$  să nu coincidă cu valoarea  $\hat{y}_1$  calculată cu formula (1) - înmulțind pe  $x_1$  cu  $b$  și adunînd apoi  $a$  - datorită faptului că dreapta de regresie nu poate trece prin toate punctele decît în cazurile de excepție cînd acestea sînt colineare. Ba mai mult, un alt individ  $j$  poate lua după  $X$  aceeași valoare ca și precedentul ( $x_1 = x_j$ ) dar după  $Y$  o alta ( $y_1 \neq y_j$ ). Or, dreapta de regresie, ca orice dreaptă, realizează o corespondență biunivocă între valorile variabilelor, adică la o abscisă corespunde o singură ordonată. Prin urmare, dacă ne plasăm în cazul cînd prin indicii  $i$  desemnăm indivizii, înseamnă că pentru orice individ din populație vom avea un triplet de valori:

$x_1$  - valoarea în raport cu  $X$ ;

$y_1$  - valoarea în raport cu  $Y$ ;

$\hat{y}_1$  - valoarea calculată din ecuația dreptei de regresie

Cum, în general, ultimele două valori nu coincid; diferențele dintre ele de forma:

$$y_1 - \hat{y}_1$$



redă mărimea abaterilor punctelor de la dreapta de regresie, respectiv eroarea care se introduce atunci când în loc de valoarea efectivă  $y_1$  lucrăm cu valoarea calculată  $\hat{y}_1$ . Atunci este evident că dreapta căutată de noi va fi cea care, pe ansamblu, va realiza cele mai mici abateri de genul precizat. Acest "pe ansamblu" nu înseamnă simplu suma diferențelor, deoarece ele fiind și pozitive și negative prin adunare se pot anula chiar la valori absolute foarte mari. Va trebui deci înlăturat întâi semnul diferențelor ceea ce se realizează prin ridicarea lor pătrat, și abia apoi se trece la însumare. Se obține deci expresia:

$$E_y = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

sau, cu expresia lui  $\hat{y}_i$  din (1)

$$E_y = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (2)$$

care reprezintă suma pătratelor abaterilor valorilor calculate de la valorile observate, pentru variabila Y.

Coeficienții  $a$  și  $b$  ai dreptei de regresie se determină de așa manieră încât mărimea  $E_y$  să fie minimă. Determinarea acestor coeficienți este întotdeauna posibilă și procedura - numită curent în literatura statistică metoda celor mai mici pătrate - conduce la o soluție unică, ceea ce vrea să spună că, oricare ar fi forma norului de puncte, există întotdeauna o dreaptă de forma (1) care satisface condiția impusă. Deci dreapta de regresie există și este unică.

Trecînd peste calculele algebrice care conduc la determinarea coeficienților  $a$  și  $b$ , voi menționa doar formulele la care se ajunge:

$$b = r \frac{\sigma_y}{\sigma_x} \quad (3)$$

$$a = \bar{y} - r \frac{\sigma_y}{\sigma_x} \bar{x} \quad (4)$$



unde simbolurile din dreapta au semnificația cunoscută: coeficientul de corelație dintre  $X$  și  $Y$  ( $r$ ), abaterile standard și valorile medii ale celor două variabile.

Să vedem acum care este interpretarea coeficienților din ecuația dreptei. Coeficientul  $a$  ne dă intersecția dreptei cu axa  $OY$  și are mai puțină importanță pentru cele ce urmează. Valoarea  $b$ , care se numește panta sau coeficientul unghiular al dreptei, arată cu câte unități se schimbă valoarea lui  $\hat{Y}$  atunci când  $X$  crește cu o unitate (e vorba de unitățile de măsură specifice fiecărei variabile). De aici ajungem și la interpretarea coeficientului de corelație care intervine în expresia lui  $b$ . Să presupunem că variem pe  $X$  nu cu o unitate, ci cu o abatere standard,  $\sigma_x$ , ceea ce va produce pentru  $\hat{Y}$  o variație  $\Delta y$ , pe care urmează să o calculăm.

Adică:

$$\hat{Y} + \Delta y = a + b(X + \sigma_x)$$

sau

$$\hat{Y} + \Delta y = a + bX + b\sigma_x$$

dar cum

$$\hat{Y} = a + bX$$

rezultă, după înlocuire și reducerea termenilor identici, că

$$\Delta y = b\sigma_x$$

Înlocuind acum pe  $b$  din formula (3), obținem

$$\Delta y = r \frac{\sigma_y}{\sigma_x} \sigma_x$$

sau, simplificând:

$$\Delta y = r \cdot \sigma_y$$

Deci variația lui  $\hat{Y}$  este egală cu  $r$  abateri standard  $\sigma_y$ .

Așadar, coeficientul de corelație  $r$  arată cu câte abateri standard proprii,  $\sigma_y$ , variază  $\hat{Y}$  atunci când  $X$  se modifică cu o



abatere standard  $\sigma_x$ , acceptînd c   intre cele dou   variabile exist   o rela  ie de dependen   linear   dat   de dreapta de regresie a lui Y asupra lui X.

S   revenim acum la m  rimea  $R_y$ . Am v  zut c   ea reprezint   suma p  tratelor distan  elor de la valorile Y empirice la cele calculate prin ecua  ia dreptei de regresie; deci, intr-un fel, ea ne d   suma p  tratelor erorilor ce se comit atunci c  nd înlocuim dependen  a statistic   a lui Y fa   de X prin dependen  a func  ional   linear  . Dac   împ  r  im pe  $R_y$  la  $n$ , num  rul indivizilor,   i scoatem radicalul (pentru c   diferen  ele s  nt la p  trat), ob  inem o m  rimenie

$$S_y = \sqrt{R_y/n}$$

numit   eroare medie p  tratic  . Semnifica  ia ei este evident  : ea ne va indica abaterea medie a valorii efective,  $y_i$ , a unui individ de la valoarea calculat  . Se poate, în plus, ar  ta c  :

$$S_y^2 = \sigma_y^2 (1 - r^2) \quad (5)$$

Se observ   imediat c   dac    $r$  este egal cu unitatea, atunci  $S_y$  este nul (deci punctele nu se abat deloc de la dreapta de regresie)   i reciproc. În situa  ia în care  $r$  este egal cu zero, din (5) rezult   imediat c  

$$S_y^2 = \sigma_y^2$$

deci eroarea este maxim  , coinciz  nd cu varianta lui Y.

Din formula (5) mai rezult   o interpretare a coeficientului de corela  ie. Sco  ind:

$$1 - r^2 = \frac{S_y^2}{\sigma_y^2} \quad (6)$$

înseamn   c   m  rimea subunitar    $1 - r^2$  ne arat   ce frac  iune din  $\sigma_y^2$  este reprezentat   de  $S_y^2$ . Cum  $S_y^2$  poate fi interpretat ca o m  sur   a dispersiei punctelor în jurul dreptei de regresie, deci ca o nedeterminare r  mas   pentru variabila Y, în ciuda rela  iei surprinse



În ecuația de regresie, atunci  $1-r^2$  reflectă proportia din nedeterminarea totală ( $\sigma_y^2$ ) rămasă încă. Sau, scoțind pe  $r^2$  din (6),

$$r^2 = \frac{\sigma_y^2 - s_y^2}{\sigma_y^2}$$

rezultă că  $r^2$  arată cu ce fracțiune se reduce varianța lui Y prin introducerea dreptei de regresie.

În termeni mai puțin riguroși (dar foarte încetățeniți în limbajul statisticii sociale, fără ca, din păcate, toată lumea să le înțeleagă cu adevărat sensul) se spune că  $r^2$  arată cât din varianța lui Y este explicată de X iar  $1-r^2$  cât din ea rămâne în contul altor factori nespecificați.

Dacă inversăm acum datele problemei și presupunem că X este funcție de Y, vom ajunge la o dreaptă de forma:

$$\hat{X} = a_1 + b_1 Y \quad (7)$$

Adică, pentru fiecare  $y_i$  va rezulta un  $\hat{x}_i$  (un X calculat) și determinarea coeficienților  $a_1$  și  $b_1$  se va face punând condiția ca expresia

$$E_X = \sum_{i=1}^n (x_i - \hat{x}_i)^2$$

sau

$$E_X = \sum_{i=1}^n (x_i - a_1 - b_1 y_i)^2 \quad (8)$$

să fie minimă.

Ecuația (7), cu coeficienții determinați din această condiție, ne dă dreapta de regresie a lui X asupra lui Y sau cu Y variabilă independentă. Cele două drepte (1) și (7) sînt, în general, distincte - adică ecuația (7) nu se deduce din (1) prin explicitarea lui X -, ele reflectînd cele două puncte de vedere după care se calculează distanțele ale căror sumă de pătrate se minimizează.



Este evident că toate considerațiile făcute în primul caz, precum și formulele date, rămân valabile și aici, cu singura precizare că trebuie schimbat locul și rolul celor două variabile.

Regresia în cazul mai multor variabile. Să presupunem, de această dată, că avem definite  $m$  variabile  $X_1, X_2, \dots, X_m$ , dintre care fie  $X_1$  variabila dependentă. Cum fiecare individ ia acum câte o valoare pentru fiecare variabilă, imaginile lor pot fi interpretate ca puncte într-un spațiu cu  $m$  dimensiuni. Ecuația de regresie se poate generaliza astfel:

$$\hat{X}_1 = a + b_2 X_2 + b_3 X_3 + \dots + b_m X_m \quad (9)$$

reprezentînd un hiperplan (generalizarea noțiunii de plan) în spațiul  $m$ -dimensional. Coeficienții  $b$  din ecuația (9) se numesc coeficienții regresiei parțiale, fiecare dintre ei arătînd cu câte unități se modifică  $X_1$ , dacă o variabilă  $X_j$  se modifică cu o unitate, celelalte fiind presupuse constante.

Pornind de la coeficienții  $b$  se pot defini așa-numiții coeficienți de corelație parțială între  $X_1$  și oricare dintre variabilele din membrul drept. De pildă, dacă luăm corelația parțială între  $X_1$  și  $X_2$ , coeficientul respectiv se va nota cu  $r_{12.3\dots m}$  și va avea semnificația unui coeficient de corelație între  $X_1$  și  $X_2$  dacă se presupune că  $X_3, X_4, \dots, X_m$  sînt menținute constante, grătînd deci legătura statistică între  $X_1$  și  $X_2$  "purificată" de intervenția celorlalte variabile explicitate.

Pentru calcularea coeficienților de corelație parțială nu este neapărat necesar să se determine ecuația de regresie. Există formule de calcul prin recurență, adică plecînd de la coeficienți de ordin inferior spre cei de ordin superior. Menționez că ordinul coeficienților de corelație parțială este dat de numărul variabilelor presupuse constante, deci, în notația coeficientului, de numărul de indici de după punct. Astfel coeficientul simplu între  $X_1$  și  $X_2$ , pe care-l notăm acum  $r_{12}$  spre a pune în evidență cele două variabile, este un coeficient de ordinul zero. Un coeficient de ordinul unu se va obține "legînd" o variabilă, de pildă  $X_3$ ; el se va nota  $r_{12.3}$  și va arăta că e vorba de co-



relația dintre  $X_1$  și  $X_2$  când  $X_3$  este constantă. Dacă mai menținem, în plus, constantă și variabila  $X_4$ , atunci vom avea o corelație parțială de ordinul doi:  $r_{12.34}$  etc.

**Observație.** În notația introdusă se desparte prin punct grupul indicilor variabilelor care se corelează de cel care precizează pe cele ce rămân constante. Este clar că ordinea în care se scriu indicii din fața punctului și de după el nu are nici o importanță; de aceea se preferă în scriere ordinea lor naturală. Adică  $r_{12.34}$  este același lucru cu  $r_{21.34}$  sau cu  $r_{21.34}$  sau cu  $r_{12.43}$  dar, evident, nu și cu  $r_{13.24}$  sau  $r_{34.12}$  etc. fiindcă în aceste din urmă cazuri am schimbat poziția indicilor în raport cu punctul.

În general, formula de recurență pomenită este următoarea:

$$r_{12.34\dots m} = \frac{r_{12.34\dots m-1} - r_{1m.34\dots m-1} r_{2m.34\dots m-1}}{\sqrt{(1 - r_{1m.34\dots m-1}^2)(1 - r_{2m.34\dots m-1}^2)}} \quad (10)$$

Cum spuneam, calculul efectiv pornește de la coeficienții de ordinul zero, adică cei obișnuiți. De pildă:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

Sau

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1 - r_{24.3}^2)}}$$

etc.

**Exemplu.** Să presupunem că avem trei variabile între care se determină următoarele corelații:

$$r_{12} = 0,60; \quad r_{13} = 0,70; \quad r_{23} = 0,80$$



și ne interesează relația dintre  $X_1$  și celelalte două variabile. Corelațiile  $r_{12}$  și  $r_{13}$  sugerează legături destul de strânse între prima variabilă și celelalte două și aproximativ de aceeași intensitate (0,60 și 0,70). Ținând constantă variabila  $X_3$  și aplicând formula de recurență se obține:

$$r_{12.3} = 0,09$$

Schimbând apoi locul variabilelor  $X_2$  și  $X_3$ , se ajunge la:

$$r_{13.2} = 0,46$$

Prin urmare, corelația dintre  $X_1$  și  $X_2$  dispăre aproape complet atunci când controlăm factorul  $X_3$  ( $r_{12.3}$  este foarte aproape de zero), pe când cea dintre  $X_1$  și  $X_3$  rămâne încă la o valoare destul de ridicată - chiar dacă mai mică decât cea inițială - când facem constant pe  $X_2$ . Din corelația inițială foarte mare dintre cele două variabile "explicative"  $X_2$  și  $X_3$ , se putea deduce, de fapt, că relațiile fiecăreia dintre aceste variabile cu  $X_1$  vor fi puternic afectate de intervenția celeilalte variabile. Iată că mărimea  $r_{12.3}$  ne arată că aproape întreaga corelație inițială de 0,60 este datorată faptului că  $X_3$  este puternic corelat cu  $X_2$  și  $X_1$ .

Un ultim indicator pe care doresc să-l mai prezint - deși mai există și alții foarte importanți - este coeficientul corelației multiple. După notațiile noastre din ecuația (9) coeficientul corelației multiple, scris  $R_{1(23...m)}$ , reprezintă corelația globală dintre variabila  $X_1$  și celelalte variabile ai căror indici sînt menționați în paranteza care apare la  $R$ . Indicatorul măsoară ea și coeficientul simplu de corelație, gradul de precizie cu care ecuația de regresie (9) reproduce valorile lui  $X_1$ . Complementar, mărimea  $1 - R^2$  reflectă partea din varianța lui  $X_1$  rămasă sau fracțiunea acestei varianțe "nexplicată" de  $X_2, X_3, \dots, X_m$ .

Coeficientul corelației multiple se poate scoate din formula de mai jos, care explicită tocmai această mărime  $1 - R^2$ :

$$1 - R_{1(23...m)}^2 = (1 - r_{12}^2)(1 - r_{13.2}^2)(1 - r_{14.23}^2) \dots (1 - r_{1m.23...m-1}^2) \quad (11)$$



De pildă, revenind la cele trei variabile între care am dat mai sus coeficienții de corelație simplă, vom aplica formula:

$$1 - R_{1(23)}^2 = (1 - r_{12}^2)(1 - r_{13.2}^2)$$

Si atunci:

$$1 - R_{1(23)}^2 = (1 - 0,36)(1 - 0,2116) = 0,5046$$

$$R_{1(23)}^2 = 1 - 0,5046 = 0,4954$$

și

$$R_{1(23)} = 0,704$$

Se observă, mai întâi, că  $R_{1(23)}$  este doar cu foarte puțin mai mare decât  $r_{13}$ , ceea ce este în concordanță cu rezultatul anterior conform căruia corelația parțială între  $X_1$  și  $X_2$ , cu  $X_3$  constant, este foarte mică; adică introducerea lui  $X_2$  în analiză nu aduce un sport sensibil în determinarea lui  $X_1$  față de ceea ce ne dă  $X_3$  singur. Pe de altă parte, se vede că, folosind expresia uzuală, din valoarea 0,4954 a lui  $R_{1(23)}^2$ , variabilele  $X_2$  și  $X_3$  "explică" și nu voi înceta să pun în ghilimele acest cuvânt - 49,54% din varianța lui  $X_1$ .

O proprietate normală a coeficientului corelației multiple care poate s-a subînțelege din cele de mai sus, dar oricum trebuie subliniată, este aceea că  $R_{1(23...m)}$  este mai mare decât oricare corelație simplă  $r_{1i}$ ,  $i = 2, 3, \dots, m$ . Din comparația lui cu aceste din urmă valori se poate intui care este, cu aproximație, ordinul de mărime a corelațiilor parțiale.

Cred că și din sumarele considerații făcute în capitolul de față se poate înțelege ce rol important are analiza de corelație în studierea fenomenelor sociale cuantificabile. Calculul corelațiilor simple ne oferă o primă imagine a intensității relației dintre factorii analizați. Cum, de obicei, raporturile de determinare dintre fenomenele sociale sînt foarte complexe, deci este greu de găsit situații în care să se studieze o relație doar între două variabile, putînd face abstracție de restul factorilor, este necesar



ca următorul pas al analizei să-l constituie calculul corelațiilor parțiale și multiplă. Analogia cu analiza multivariată a caracteristicilor calitative este evidentă; ambele proceduri au rolul de a suplini metoda experimentală efectivă, căci aceste analize de raporturi în care anumiți factori se mențin constanți nu fac decât să reproducă schema logică a experimentului.

Totuși, atrag din nou atenția că ne aflăm pe tărîmul statisticii, că indicatorii calculați aici nu sînt decît niște instrumente prin care putem verifica (mai corect, infirma) anumite ipoteze teoretice privind determinațiile dintre fenomene sau, cel mult, sugera unele legături între ele. De exemplu, oricît ar fi de bogat lotul de variabile auxiliare luat pentru a preciza relația dintre  $X_1$  și  $X_2$ , nu vom fi niciodată siguri că indicele de corelație parțială  $r_{12.34...m}$  nu se modifică radical prin introducerea unei noi variabile, omisă pînă atunci. Rămîne sarcina teoriei sociologice de a analiza conținutul variabilelor manipulate, de a stabili schemele explicative necesare.



## PARTEA A IV-A

### PROBLEMATICA EȘANTIONĂRII

#### Cap. XI. PROCEDEE DE EȘANTIONARE

##### 1. Principiile cercetării selective

După cum am precizat în mai multe rânduri, în cercetările sociologice sîntem, de cele mai multe ori, în situația de a investiga proprietățile, însușirile etc. ale unui număr mare de unități (obiecte-suport, cum le-am numit), adică avem de a face cu o abordare cantitativă la nivelul obiectului, ceea ce duce la folosirea instrumentelor statistice. Este de subliniat că, în toate aceste cazuri, indiferent de natura obiectelor purtătoare de însușiri, prima operație ce trebuie realizată, pentru a putea desfășura în bune condițiuni cercetarea, este delimitarea, circumscrierea cît mai exactă a populației, respectiv a universului de obiecte asupra cărora cercetătorul își îndreaptă atenția. Am precizat în capitolul referitor la etapele cercetării atît importanța unei delimitări corecte a populației, cît și dificultățile practice legate de această operație și nu mai revin asupra lor. Ceea ce ne interesează în momentul de față este doar faptul că înaintea oricărei cercetări efective de teren trebuie ca cercetătorul să posede o delimitare a populației ce va fi investigată și asupra căreia să fie valabile concluziile ce rezultă în urma cercetării.

În situația în care toți indivizii din populația astfel determinată sînt direct investigați, cercetarea se numește complet. Dacă însă populația este foarte mare, depășind posibilitățile sociologului de a o investiga exhaustiv într-un timp util, acesta se va mulțumi să intre în contact direct doar cu o parte din universul de indivizi și atunci cercetarea se numește selectivă. Partea din populație aleasă spre investigare va fi numită eșantion, iar totalitatea operațiilor prin care se construiește eșantionul



poartă numele de egantionare sau selectie.

Chiar în acest punct inițial al expunerii simt nevoia să fac o precizare de maximă importanță. A efectua o cercetare selectivă nu înseamnă a reduce populația sau universul cercetării; ea înseamnă numai că informația referitoare la populație ÎN TOTALITATEA SA nu se recoltează de la toți indivizii, ci numai de la o parte dintre ei. Principiul gnoseologic care stă la baza cercetărilor selective este ideea că întotdeauna o parte reflectă, într-o anumită măsură, întregul. Principiul practic ce ghidează activitatea propriu-zisă de egantionare îl constituie maximizarea măsurii în care partea reflectă întregul.

Nici ideea, nici practica egantionării nu sînt lucruri noi. Ele există, s-ar putea spune, de cînd există activitate umană și nu sînt, prin urmare, caracteristice doar cercetărilor sociologice. Egantionări se efectuează atît în cadrul diferitelor cercetări de natură științifică, cît și în activitatea curentă, de zi cu zi, a oamenilor. O gospodină care gustă supă spre a vedea cît e de sărată, un geolog care ia o mostră de roci dintr-un munte sau astronautul de pe Lună, asistenta medicală ce recoltează o cantitate mică de sînge, controlorii tehnici de calitate ce analizează amănunțit o parte din piesele produse într-o uzină etc. toți aceștia efectuează, în fond, cercetări selective.

După cum se vede și din exemplele date, uneori cercetarea selectivă este impusă de realitate, fiind imposibil un studiu complet (gospodina ar trebui să mănînce întreaga oală cu supă, astronautul să aducă Luna pe Pămînt etc.) Dar și în cazurile în care este teoretic posibilă studiarea completă a populației, egantionarea își dezvăluie o serie de avantaje incontestabile, care o fac foarte larg răspîndită în toate domeniile ce utilizează analiza statistică.

Sintetizînd avantajele aduse de cercetările selective în raport cu cele complete, cel puțin pentru domeniul sociologiei, se poate afirma că acestea apar pe următoarele paliere principale:

a) Avantaje de ordin practic-economic. Studiile selective se realizează, evident, la un cost financiar mai scăzut decît cele complete, fac apel la un număr mai mic de persoane calificate sau auxiliare de cercetare, realizîndu-se deci și o economie de muncă.



b) Avantaje de ordin aplicativ. Rezultă, în principal, din scurtarea timpului necesar realizării cercetării, ceea ce-i conferă acesteia posibilitatea de a oferi soluții prompte. Asemenea gen de avantaje este extrem de important mai ales în domeniul științelor sociale, unde schimbările se produc cu mare rapiditate și nu o dată rezultatele cercetărilor își pierd actualitatea până în momentul obținerii sau publicării lor. Așa de pildă, datele recensămintelor (exemplu clasic de cercetare socială completă) apar tipărite după 2-3 ani sau chiar mai mult de la data efectuării lor, când ele capătă doar, ca să nu zic valoare istorică, o semnificație de punct de referință pentru datele demosociale; dar este evident că utilitatea lor ar fi cu totul alta dacă cercetătorii și organele decidente le-ar avea la dispoziție mai rapid.

c) Avantaje de cunoaștere. Dacă cele spuse mai sus sînt chestiuni foarte simple și evidente, avantajele de cunoaștere ale studiilor selective par mai neobișnuite, avînd în vedere că, în principiu cercetarea unui eșantion nu poate să dea decît o informație incompletă, mai săracă decît cea rezultată din studiile exhaustive. Și totuși, nu rareori, în practică, lucrurile se inversează. Intr-adevăr, studierea unui lot mai restrîns de indivizi face posibilă utilizarea unui personal auxiliar (de exemplu, operatori de anchetă) mai puțin numeros, ce poate fi temeinic instruit, verificat (din punctul de vedere al aptitudinilor necesare cercetării, al trăsăturilor morale și de personalitate etc.) și controlat în activitatea concretă de teren. Prin urmare, se pot reduce substanțial erorile ce apar în procesul de culegere a datelor, erori care, în fapt, ocupă, de regulă, locul cel mai important în rîndul factorilor "perturbatori" ai oricărei cercetări sociologice.

Pe de altă parte, numărul mai mic de indivizi investigați dă posibilitatea, ca în aceleași condiții de timp și de cheltuieli, să se realizeze o cercetare de mai mare adîncime. Și timpul dar și personalul restrîns și bine instruit oferă șanse de a se aborda aspecte mai complexe, de a se aplica tehnici mai rafinate și deci de a se ajunge la rezultate mai consistente, de esență, atunci cînd investigăm doar un eșantion. În plus, prin posibilitatea de a aplica



tehnică diferite aceluiși lot de indivizi, rezultatele obținute pot fi controlate, coroborate și astfel se pot elimina încă o serie de erori. Una este, de exemplu, informația oferită de o anchetă ce folosește un chestionar cu 20-25 de întrebări foarte simple, cum se întâmplă când populația e mare și se face un <sup>exhaustiv</sup> studiu al ei, și alta cea dată de un chestionar bogat, aplicat pe îndelete, capabil să surprindă mult mai multe aspecte, din corelarea cărora se pot degaja concluzii de mai mare finețe.

În sfârșit, tot pe linia unor asemenea avantaje se află și cele legate de prelucrarea datelor în timp optim. Este limpede că, într-un același interval de timp, prelucrarea unui număr imens de chestionare, să zicem, nu poate să includă decât operații simple de tabulare, pe când dacă volumul e mai redus și prelucrarea poate fi mai rafinată, timpul și fondurile materiale la dispoziție dând posibilitatea unor operații și analize statistice suplimentare. Desigur că însăși pregătirea pentru prelucrare (verificarea chestionarelor, postcodificarea etc.) se poate face în condiții mult mai bune când numărul acestora este mai redus.

Pentru realizarea operațiilor practice de eșantionare, se ridică trei probleme de bază legate de:

- a) mărimea eșantionului
- b) reprezentativitatea acestuia și
- c) modul de alegere a elementelor constitutive.

Toate aceste trei aspecte sînt strîns intercorelate și este dificilă o analiză separată a lor. Voi începe cu o tratare sumară a primelor două, - asupra cărora voi reveni cu determinări mai precise în capitolul următor -, urmînd apoi să descriu mai pe larg pe cel de al treilea, în concordanță deci cu titlul acordat capitolului de față.

Să luăm ca punct de plecare mărimea eșantionului și să vedem care sînt factorii ce concurează la alegerea unui eșantion cu o dimensiune dată. Cei mai importanți par a fi următorii patru:

(1) Caracteristicile populației Așa cum am mai spus, asupra unei populații se studiază seturi de însușiri, caracteristici etc. În funcție de acestea populația noastră poate manifesta un grad mai mare sau mai mic de omogenitate. Pentru studierea caracteristicilor față de care populația este omogenă, va fi suficient un eșantion de mărime mică. Ca să reiau un exemplu an-



terior, gospodina nu are nevoie decât de o linguriță de supă spre a-și da seama dacă e sau nu suficient de sărată (însușirea ce se analizează), dat fiind că întreg conținutul oalei este omogen și bine "amestecat" și nu este de așteptat ca într-o parte supa să fie mai sărată și în alta mai puțin. Grupurile sociale evident că nu se caracterizează printr-o asemenea omogenitate, chiar în cazul unor variabile foarte simple. Și nu numai că eterogenitatea este, în general, mare, dar populația nu este nici atât de bine amestecată ca supra în oală, căci, așa cum se știe, există o clară tendință ca tipurile asemănătoare să se grupeze (Moser). Acest lucru ridică mult riscul de a trage concluzii pe baza unui eșantion restrâns, care n-ar fi capabil să acopere întreaga arie a unor astfel de grupări.

Cum în cercetările efective de teren se urmăresc întotdeauna mai multe caracteristici, va trebui să se ia în considerare pentru a stabili mărimea eșantionului aceea care are gradul de dispersie maxim. În cazul caracteristicilor cantitative, așa cum s-a văzut în capitolele anterioare, dispersia poate fi măsurată prin indicatori numerici - în speță, abaterea standard - indicatori care vor fi legați direct proporțional de mărimea eșantionului.

(ii) Gradul de detaliere a analizei. Este vorba aici de un lucru simplu și evident, dar care nu este luat întotdeauna în considerare. Un volum al eșantionului poate fi suficient de mare pentru descrierea unei caracteristici în ansamblu, dar nu și pentru realizarea unor comparații între situațiile unor subeșantioane, delimitate cu o altă caracteristică. De exemplu, vom avea nevoie de un eșantion pentru a studia, să spunem, numărul mediu de camere ce revin unei familii din municipiul Cluj-Napoca și de un altul, mult mai mare, dacă urmărim să estimăm aceeași mărime dar la nivelul cartierelor orașului.

Aici se cuvine făcută o precizare importantă a cărei semnificație - ca de altfel și a altor afirmații de mai înainte - se va dezvălui treptat pe parcursul acestui capitol și al următorului. Nu există o rețetă universală pentru a spune cât de mare trebuie să fie un eșantion și, mai ales, nu există o rețetă care să ne dea mărimea eșantionului ca procent din numărul indivizilor ce compun populația.



Mai întâi că - și cei cu o oarecare perspicacitate, dar fără cunoștințe statistice, vor sesiza-o cu surprindere - printre factorii principali ce determină mărimea eșantionului nu intervine mărimea populației. Acest lucru poate părea paradoxal, dar el primește o demonstrație riguros matematică. (De fapt, mărimea populației influențează pe cea a eșantionului doar într-un caz mai rar întâlnit în sociologie, când cele două dimensiuni sînt apropiate. Dacă se întîlnesc totuși asemenea cazuri, e recomandabil să se studieze populația exhaustiv, căci întreg arsenalul teoretico-matematic al eșantionării pornește de la principiul că se lucrează cu populații mai.) În al doilea rînd, că, din punct de vedere al problematicei eșantionării, egalitatea unor mărimi relative nu este totdeauna operantă. Mai exact, un eșantion de 5%, de exemplu, dintr-o populație de 10.000 unități nu are nici pe departe aceeași putere de reflectare a întregului ca un eșantion tot de 5% dar dintr-o populație de 1.000.000 de unități.

Am făcut aici aceste precizări, legate chiar de exemplul evocat mai sus, pentru că efectiv mi s-a întîmplat ca un beneficiar să nu poată înțelege că un eșantion bun pentru municipiul Cluj, în ansamblu, nu mai este bun pentru un cartier oarecare, deși din întreg orașul și din fiecare cartier al său a intrat în eșantion exact același procent din familie.

(iii) Reprezentativitatea eșantionului. Este probabil noțiunea fundamentală în cercetările selective și de aceea ar fi fost poate mai normal să încep cu ea sau chiar ca întreg setul acestor considerații să se raporteze la ea și nu la dimensiunea eșantionului. Am ales totuși această din urmă noțiune ca "punct de sprijin" pentru motivul că e mai ușor de înțeles și pentru că reprezentativitatea va primi o definiție riguroasă, sub formă matematică, doar în capitolul ce urmează. Pentru moment, vom spune că reprezentativitatea unui eșantion este gradul de concordanță între caracteristicile definite pe acesta (ca parte) și cele similare definite la nivelul întregii populații (ca întreg). Deci reprezentativitatea este, altfel spus, măsura în care partea reflectă întregul.

Chiar și din această definiție imprecisă rezultă un



lucru fundamental, pe care puțină lume îl cunoaște și îl înțelege cu adevărat: nu există eșantionare reprezentative și nere-representative, ci există doar eșantioane mai mult sau mai puțin reprezentative pentru o populație și o caracteristică a acesteia date. Orice eșantion posedă o reprezentativitate diferită de zero și nici un eșantion nu este 100% reprezentativ sau, mai corect spus, probabilitatea de a se întâmpla acest din urmă lucru este practic nulă.

Între mărimea eșantionului și reprezentativitatea sa există desigur o dependență directă, în sensul că, toți ceilalți factori fiind egali, pe măsură ce crește volumul eșantionului sporește și gradul său de reprezentativitate, reciproc, pentru a spori reprezentativitatea, în condiții egale, trebuie mărit volumul eșantionului. Un eșantion de mărime nulă - dacă se poate vorbi de așa ceva - va avea o reprezentativitate nulă și unul egal cu întreaga populație va avea una completă (să-i zicem 100%). Dar această dependență este departe de a fi lineară, deci creșterile pe una din cele două dimensiuni menționate nu antrenează creșteri proporționale pe cealaltă. Spre a sugera cât de eșt forma dependenței dintre cele două mărimi, am construit graficul din figura 1, în care se vede că la o creștere a valorilor mici ale mărimii eșantionului corespund creșteri rapide ale reprezentativității; dimpotrivă, după ce numărul de indivizi depășește o anumită zonă, marcată de segmentul AB, practic orice spor al acestuia aduce o creștere infirmă a reprezentativității. Prin urmare, câștigurile realizate de pe urma mărimii exagerate a eșantionului sînt minore și ele, de regulă, sînt contrabalansate de creșterea gradului de eroare implicat de o mulțime de factori atunci cînd cîmpul de investigație se lărgeste (și asta fără a mai vorbi de ridicarea costului cercetării și de alte impedimente practice.).



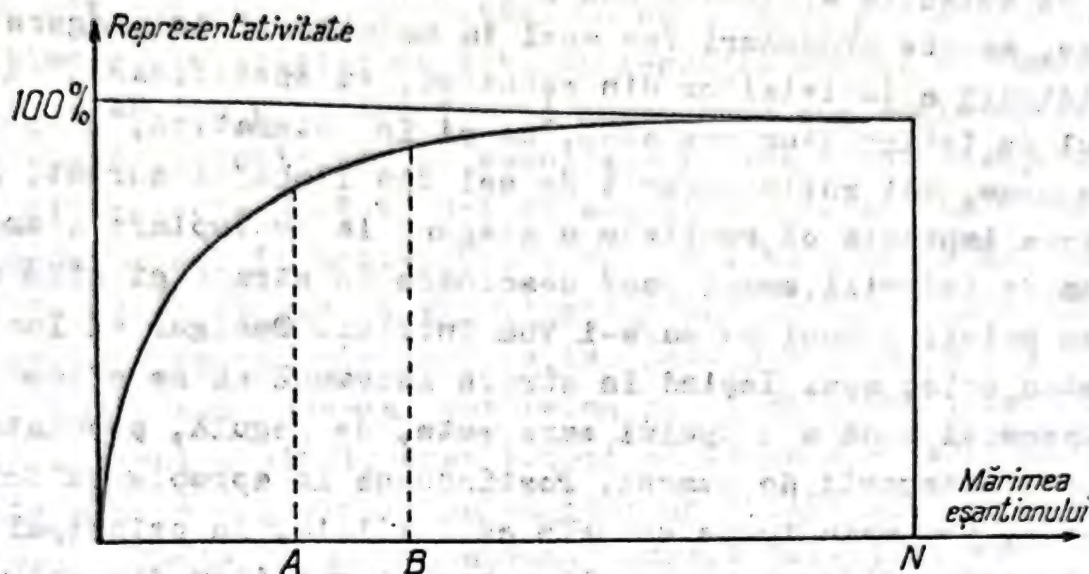


Fig. 1 Relația dintre volumul eșantionului și reprezentativitatea sa ( $N$  = volumul populației)

(iv) Procedeul de eșantionare. Dat fiind că nivelul de reprezentativitate al eșantionului (atunci când el poate fi calculat) este în funcție de tehnica de selecție, este clar că și mărimea eșantionului depinde de acest factor. La un același nivel de reprezentativitate dat, folosind procedee de eșantionare diferite, vom avea nevoie de eșantioane de mărimi diferite.

## 2. Tipuri de eșantioane

Una dintre problemele de bază ale cercetării selective este aceea a procedeele practice prin care se aleg din populație elementele constitutive ale eșantionului ce urmează investiga. În principiu, toate lucrările metodologice admit existența a două mari clase (tipuri sau genuri) de eșantioane: cele alese printr-o procedură probabilistă, numite eșantioane aleatoare și cele care nu fac apel la astfel de proceduri, eșantioane nealea-



toare.

Primul lucru asupra căruia ne vom opri este sensul noțiunii de selecție aleatoare sau probabilistă. După cum se poate înțelege, aceste proceduri fac apel la modalități de alegere întâmplătoare a indivizilor din eșantion, cu specificarea că termenul de întâmplător are aici, ca și în matematică, un sens, dacă nu opus, cel puțin diferit de cel din limbajul curent. Am putea avea impresia că realizăm o alegere la întâmplare a unui eșantion de indivizi umani dacă descindem în stradă și stăm de vorbă cu primii oameni pe care-i vom întâlni. Desigur că lucrurile nu stau deloc așa. Ieșind în stradă înseamnă că ne aflăm într-o anumită zonă a orașului care este, de regulă, populată cu anumite categorii de oameni. Postîndu-ne în apropierea Universității, vom avea toate șansele să întâlnim, în principal, studenți, cadre didactice sau alt personal muncitor din Universitate. Similar se petrec lucrurile dacă locul nostru de descindere este o zonă industrială, comercială etc. Apoi, la o anumită oră din zi, șansele sînt foarte mari să întâlnim sau să nu întâlnim pe stradă anumite categorii de oameni. Nu mai vorbesc de faptul că fiecare dintre noi, oricîtă obiectivitate și-ar impune, are tendința să aleagă și să respingă anumite tipuri de persoane. Așadar, este limpede că o asemenea alegere "întâmplătoare" nu poate satisface cerințele unei cercetări selective serioase, deoarece eșantionul este departe de a surprinde structurile de bază ale populației și deci este de așteptat că și în alte privințe, referitor la alte caracteristici, chiar dacă e vorba de opinii, motivații etc., el nu va da o imagine suficient de corectă a realității.

Întîmplarea matematică trebuie căutată prin procedee riguroase, care să elimine orice alegere subiectivă sau orice intervenție a unui factor perturbator cu acțiune constantă (cum este abordarea oamenilor pe stradă). În general, se apreciază că pentru realizarea unui eșantion probabilist este nevoie ca fiecare element din populație să aibă o șansă CALCULABILĂ și NONNULĂ de a fi ales. Orice eșantionare care nu respectă aceste cerințe va fi considerată nealeatoare.

Diferențierea conceptuală a celor două tipuri de selecție este foarte importantă, chiar dacă, din punctul de vedere al



eficienței practice, este imposibil să se indice vreo superioritate absolută a unui gen sau a altuia. Vorbind de superioritate, mă refer desigur la posibilitatea pe care o are eșantionul de a reflecta cât mai corect populația. În principiu, poate fi ales un eșantion probabilistic după toate regulile artei, dar care, întâmplător să dea o eroare foarte mare, după cum poate fi studiat un eșantion nealeator cu rezultate foarte bune. Esența deosebiri dintre cele două tipuri nu constă nici măcar în faptul că, în general sau în medie, unele sînt mai bune decît altele, ci doar în aceea că pentru eșantioane probabiliste se poate calcula gradul de reprezentativitate, iar pentru celelalte nu.

Deși, cum spuneam, voi relua problema reprezentativității în ultimul capitol, cred totuși că trebuie să arăt aici ce se înțelege prin calculabilitatea reprezentativității, lucru fără înțelegerea căruia nu se pot aprecia corect virtuțile eșantionării aleatoare. Să presupunem că prin cercetarea selectivă urmărim să estimăm o valoare  $y$  a unei caracteristici din populație (de exemplu, vîrsta medie, retribuția medie, procentul celor ce exprimă o opinie etc.). Cercetînd efectiv indivizii din eșantion, vom da peste o valoare corespunzătoare  $y'$ , care, în general, este diferită de  $y$ . Este clar că un eșantion este cu atît mai reprezentativ cu cît  $y'$  se apropie mai mult de  $y$ . S-ar părea că acea calculabilitate de care pomeneam n-ar consta în altceva decît în efectuarea diferenței dintre  $y$  și  $y'$  (abaterea lui  $y'$  de la  $y$ ). Numai că  $y$  este o mărime necunoscută și prin urmare este imposibil să se calculeze această diferență. Cea ce oferă eșantionarea aleatoare este posibilitatea de a estima această diferență în termeni de probabilități. Mai exact, pentru orice eșantion aleator, se poate afirma cu o anumită probabilitate  $P$ , calculabilă, că mărimea din eșantion,  $y'$ , nu diferă de cea adevărată (din populație),  $y$ , cu mai mult decît o mărime  $\Delta$ , ce poate fi făcută oricît de mică, așa cum probabilitatea  $P$  poate fi făcută oricît de mare.

Să trecem în continuare, la prezentarea pe scurt a principalelor tipuri de eșantioane, începînd cu cele probabiliste. Spun pe scurt deoarece problematica acestui paragraf este extrem de vastă și nu-i voi putea face decît un succint rezumat. Pentru mai multe detalii, fac trimitere la lucrarea lui C.A.Moser (1967)



tradusă în românește, unde se pot găsi informații și precizări suplimentare, dezvăluind, în același timp, și rolul important pe care această carte l-a avut în organizarea ideilor din capitolul de față.

## 2.1. Eșantionarea aleatoare

Tehnicile de selecție aleatoare, sînt la rîndul lor, clasificate în mai multe tipuri principale ce se aleg în funcție de necesitățile studiului, de răspîndirea - respectiv, gruparea - populației în teritoriu, de baza informațională existentă înainte de începerea cercetării etc. Si dacă am pomenit de baza informațională, să mai fac o precizare importantă: o procedură de eşantionare aleatoare - în sensul cel mai riguros al termenului - nu poate fi utilizată decît atunci cînd există un cadru de eşantionare corect întocmit. Prin cadru de eşantionare se înțelege o listă completă a populației din care să se extragă eşantionul sau o altă informație despre populație echivalentă cu aceasta (de exemplu, o hartă pe care să fie marcate toate unitățile). Se înțelege că întocmirea unui cadru de eşantionare presupune, pe lîngă o definiție operațională riguroasă a populației, și existența unor posibilități practice pentru cercetător de a identifica efectiv (și nu potențial sau în principiu) fiecare element al populației. Lucrul acesta nu este deloc simplu și dacă sociologul nu posedă dinainte un asemenea cadru de eşantionare, din stocul de date existente la anumite organe sau instituții (consilii populare, miliție, poștă etc.) este foarte puțin probabil ca el însuși să aibă posibilitatea de a și-l construi.

Procedeele de bază utilizate în eşantionarea aleatoare sînt:

- A. Procedee elementare (eşantionarea simplă aleatoare)
- B. Eşantionarea prin stratificare
- C. Eşantionarea multistadială
- D. Eşantionarea multifazică

Să trecem la prezentarea lor în această ordine.

A. Procedee elementare. Cel mai simplu, ca principiu, procedeu elementar este așa-numita metodă a loteriei: indivizii populației care vor constitui eşantionul sînt "trăși la soți" cu



ajutorul unei urne, unde fiecare unitate este prezentă, de exemplu sub forma unui bilețel sau a unei bile. Toată lumea cunoaște precauțiunile ce se iau la astfel de extrageri (cum ar fi cele de la loto sau pronoeexpres): bilele trebuie să aibă dimensiuni și mase identice, urna trebuie să permită o cât mai bună amestecare a lor etc. Toate aceste riguri sînt impuse de obiectivul fundamental - care rămîne valabil și în egantionarea în scopuri științifice -, și anume acela de a asigura tuturor unităților șanse egale de a fi extrase.

Dacă populația este formată din  $N$  indivizi, atunci este evident că, în condiții ideale, probabilitatea ca un individ să fie extras este  $1/N$ . Se observă, că după extragerea unui individ, probabilitățile pentru ceilalți se schimbă puțin: înainte de extragerea celui de al doilea în urmă se află  $N-1$  indivizi și deci probabilitatea devine  $1/(N-1)$ ; înainte de extragerea celui de al treilea, probabilitatea va fi  $1/(N-2)$  ș.a.m.d. Deci șansele de a fi prezenți în egantion, pentru indivizii rămași în urnă, cresc pe măsură ce se înaintează cu aceste extrageri. Pentru a se menține mereu aceeași probabilitate, se poate proceda la reintroducerea în urnă a fiecărui individ extras, totalul fiind atunci mereu  $N$ . În felul acesta se face distincția între două tipuri de selecții: repetată (cînd se reintroduc în urnă indivizii extrași) și nerepetată. Diferența nu are o importanță practică deosebită decît atunci cînd volumul populației este relativ mic, adică atunci cînd scăderea unui număr de indivizi din total (număr ce merge pînă la volumul egantionului minus unu) influențează sensibil probabilitățile  $1/(N-i)$ . Teoretic, diferența între cele două proceduri își pune amprenta asupra formulelor de calcul ce intervin.

Utilizarea efectivă a metodei loteriei este foarte restrînsă, în cazul cercetărilor sociologice de teren. Motivul fundamental rezidă în numărul mare de indivizi ce formează, de regulă, populația și pentru care este imposibil să se asigure condițiile tehnice necesare unei "trageri la sorți" riguroase. A scrie mii, zeci sau sute de mii de bilețele perfect egale, a le amesteca suficient de bine pentru a elimina orice efect al factorilor ce au determinat ordinea introducerii lor în urnă, a le extrage apoi din urnă cu totul întîmplător sînt cerințe foar-



te greu de îndeplinit. Din acest motiv se caută substituirea acestor procedee cu altele, echivalente sau aproape.

Unul dintre acestea, de o mare rigurozitate și, în același timp, cu o eficiență practică deosebită constă în utilizarea tabelelor cu numere aleatoare. Așa cum o arată și numele e vorba de tabele ce conțin o mulțime de numere determinate de o manieră strict probabilistică. Trece peste detaliile de construcție a lor, dar precizez că se utilizează metode foarte sofisticate și nu este pur și simplu o înșiruire de numere așa cum îi vin cuiva în minte. Numerele pot fi formate din 4,5,6 sau mai multe cifre și sînt așezate, de regulă, sub forma unor coloane în felul următor:

452402	390946	...
419415	886954	...
963827	250162	...
710612	748522	...
981450	054556	...
. . . . .		

Modul de utilizare a tabelelor este foarte simplu. După cum am spus, pentru orice eșantionare aleatoare posedăm sau trebuie să construim o listă a tuturor unităților din populație, listă pe care indivizii pot fi numerotați - și de regulă, numerotarea se face chiar odată cu întocmirea listei - de la 1 la  $N$ . Alegerea unui eșantion de volum  $n$  revine la următoarele: se pornește dintr-o zonă oarecare a tabelului de numere aleatoare, pe coloane sau pe linii, și numerele întîlnite ne vor indica poziția indivizilor din lista noastră ce vor intra în eșantion. Dacă tabelul conține numere cu mai multe cifre decât are  $N$ , atunci se renunță la primele cifre ale numerelor din tabel. Dar mai bine să recurgem la un exemplu concret. Să presupunem că populația noastră este formată din 7200 indivizi, din care vrem să selecționăm 600. Folosind un tabel similar celui din care am extras numerele de mai sus, se vede că va trebui să renunțăm la primele două cifre, iar dintre numerele cu patru cifre rămase, de asemenea trebuie eliminate cele mai mari decât 7200. Pornind chiar de la primul număr scris pe coloane, primul individ extras va fi cel cu numărul 2402, al doilea cel cu numărul 3827 (s-a sărit peste



9415 care depășește valoarea lui  $N$ ), al patrulea 612 (zeroul din față nu se ia în considerare), al cincilea 1450 ș.a.m.d. pînă se completează suma de 600.

O formă de eșantionare - deseori folosită și considerată a poseda atributul de aleatoare - este așa-numita metodă a pasului. Reluînd exemplul anterior, cu presupunerea că trebuie să alegem 600 indivizi din 7200, rezultă, din împărțirea:

$$7200 : 600 = 12$$

că tot unul din 12 indivizi va fi ales în eșantion. Se ia atunci lista celor 7200 de unități, se alege la întâmplare un individ din primii 12 aflați pe listă (de pildă prin tragere la sorți) după care numerele de ordine ale celorlalți se vor obține adăugînd succesiv cifra 12 la numărul anterior determinat. Astfel, dacă am ales în eșantion pe al 5-lea de pe listă, următorul va purta numărul de ordine 17(5+12) etc. La sfîrșit vom obține 600 de indivizi exact cît ne-am propus.

Valoarea

$$k = \frac{N}{n}$$

se numește pasul de eșantionare. Raportul poate duce la o valoare întreagă, așa cum s-a întîmplat în cazul ilustrat, și astfel procedura ne dă efectiv un eșantion de mărimea urmărită. Dacă pasul  $k$  iese o mărime fracționară, atunci se poate lucra cu întregul imediat mai mic decît  $k$  - partea întreagă a lui  $k$  - sau cu cel imediat superior - partea întreagă a lui  $k+1$ . În prima situație, volumul eșantionului va depăși cifra prevăzută, ceea ce înseamnă că dintre indivizii aleși o parte trebuie eliminați. Acest lucru se poate realiza folosind tot metoda pasului selectînd din eșantionul rezultat în prima fază numărul suplimentar de indivizi. În al doilea caz, volumul eșantionului va fi mai mic decît cel prevăzut și numărul indivizilor necesari spre completare se poate determina printr-o nouă selecție asupra întregii populații, cu un pas corespunzător.

Metoda pasului nu este stricte sensu una aleatoare, deoarece odată cu alegerea primului individ (să zicem al  $i$ -lea) se determină exact și ceilalți, avînd rangurile  $i+k$ ,  $i+2k$ ,  $i+3k$



etc. Deci pentru restul probabilitatea de a fi aleși este nulă, ceea ce contrazice principiul însuși al eșantionării aleatoare. Dar, în practică, metoda dă rezultate bune, mai ales atunci când listarea se face după un criteriu (de pildă cel alfabetic) ce nu are nici o legătură cu probleme ce urmează a fi studiate. În caz contrar, se pot produce unele erori sistematice. De exemplu, dacă cele 7200 unități sînt localități pentru care se urmărește estimarea mărimii medii și dacă ele sînt înșirate în ordinea crescătoare a volumului lor demografic, atunci cînd elementele alese în eșantion poartă numere de ordine mici - de pildă, 1, 13, 25 etc. - se va subestima mărimea cercetată, iar cînd poartă numere mari (12, 24, 36 etc.) se va supraestima.

B. Eșantionarea prin stratificare. Este o procedură care dă rezultate mai bune în privința reprezentativității decît cea simplă aleatoare. Mai exact, la același volum al eșantionului, nivelul de reprezentativitate este mai ridicat în cazul eșantionării prin stratificare. Ideea de bază constă aici în divizarea populației în straturi, în funcție de una sau mai multe caracteristici, pentru fiecare strat realizîndu-se separat o eșantionare aleatoare. În felul acesta, mărimile subeșantioanelor corespunzînd diferitelor straturi vor fi proporționale cu ponderea acestor straturi în populația totală, avînd garanția că fiecare strat este fidel reprezentat în eșantion. Se limitează așadar jocul șansei, fără ca prin aceasta să se încalce principiile eșantionării aleatoare.

Drept factori de stratificare se aleg, de preferință, acele variabile care au o influență importantă asupra fenomenelor studiate. Dacă este vorba de un univers de indivizi umani, asemenea factori vor fi: sexul, vîrsta, ocupația, școlaritatea, nivelul veniturilor etc. Un rol similar îl poate juca apartenența la unități administrativ-teritoriale, la diferite grupuri sau instituții sociale.

Să dăm și un exemplu, pentru a înțelege mai clar semnificația și modalitatea practică de realizare a procedurii. Să presupunem că dintr-o populație de persoane, în mărime de 100.000, ocupate în industria unui județ, dorim să investigăm direct un eșantion de 1.000 indivizi. Grupînd persoanele ocupate pe categorii socio-profesionale, vom ajunge, să zicem, la situa-



ția de mai jos:

muncitori	- 85.000
maistri	- 7.500
tehnicieni	- 3.500
funcționari	- 2.500
ingineri, econo- miști	- 1.500

---

Total N = 100.000

Drept urmare, eșantionul nostru de 1.000 de persoane va trebui să cuprindă unu la sută din fiecare categorie, adică: 850 muncitori, 75 maistri, 35 tehnicieni, 25 funcționari și 15 ingineri și economiști, reflectând exact ponderea fiecărei ocupații în populație.

De multe ori aceste stratificări au drept scop nu doar o reprezentare corectă a straturilor, ci realizarea unui studiu comparativ privind situația fiecăruia dintre straturi. În astfel de cazuri, procedura nu mai poate fi utilizată exact în forma simplă ilustrată mai sus, căci anumite straturi de dimensiune redusă în populație vor fi numerie slab reprezentate în eșantion. Revenim, de fapt, la o problemă atinsă și în paragraful precedent, care, de data aceasta, constă în faptul că dacă cei 850 de muncitori pot constitui un eșantion cu o reprezentativitate suficientă în raport cu populația muncitorilor, pentru restul categoriilor eșantioanele nu mai posedă această proprietate, deși, ca pondere ele sînt egale (1%). În cazuri ca acesta, nu se vor putea face comparații între straturi, iar pentru realizarea unui asemenea obiectiv vor trebui mărite simțitor subeșantioanele de mică dimensiune. Dar atunci nu va mai apare corect imaginea globală, ea fiind evident deformată de suprareprezentarea unor straturi.

Pentru a împăca ambele cerințe, se poate realiza studiul cu un eșantion deformat, în care fiecare subpopulație să fie reprezentată de un număr de indivizi corespunzător unei reprezentativități dorite pe subpopulații, iar apoi, la constituirea imaginii de ansamblu, mărimile estimate pe straturi se



vor pondera în funcție de mărimea reală a stratului. Astfel, dacă în exemplul de mai sus, estimând media câștigurilor lunare pentru cele 5 categorii - din studiul unor eșantioane suficiente de reprezentative pentru fiecare - obținem valorile  $\bar{x}_1$ ,  $\bar{x}_2$ ,  $\bar{x}_3$ ,  $\bar{x}_4$  și  $\bar{x}_5$ , media estimată pe întreaga colectivitate se poate calcula în felul următor:

$$\bar{x} = 0,850\bar{x}_1 + 0,075\bar{x}_2 + 0,035\bar{x}_3 + 0,025\bar{x}_4 + 0,15\bar{x}_5$$

Mai există, desigur, și o altă posibilitate de ajustare a datelor globale. După studierea subeșantioanelor pentru fiecare strat se pot alege eșantioane mai mici, proporționale cu talia stratului, care să se reunească pentru a obține un eșantion fidel ansamblului. Studiul acestui eșantion va da estimările pentru întreaga populație. Preferabilă pare a fi prima metodă căci aceasta din urmă poate introduce erori suplimentare la selecția a doua.

În fine, o observație ce se impune în legătură cu eșantionarea prin stratificare. Este clar că avantajele sale ar trebui să conducă la o largă utilizare a procedurii, numai că în realitate lucrurile stau altfel. Condiția fundamentală - cunoașterea structurilor populației înainte de studierea ei - este rareori îndeplinită; statisticile noastre curente dau puține informații care să poate fi utilizate de cercetarea sociologică și asta nu numai pentru construirea unor eșantioane de înaltă precizie.

C. Eșantionarea multistadială. Se folosește probabil mai des decât cea precedentă dar nu din rațiuni de ameliorare a estimărilor, ci din altele, de natură practică. Se știe că populațiile statistice nu sînt distribuite uniform în teritoriu sau în alt spațiu de definire al lor. Ele au tendința de a se concentra sub formă de grupuri mai mult sau mai puțin delimitate. Dacă e vorba de indivizi umani, cum se întîmplă de multe ori în studiile noastre, aceștia aparțin unor grupuri familiale, rezidențiale, de muncă sau învățătură etc. La rîndul lor, aceste grupuri de gradul întâi, să le zicem, se pot diviza în grupuri secundare, care la rîndul lor ... etc.

Principiul eșantionării multistadiale constă într-o selecție indirectă a indivizilor ce formează eșantionul, prin in-



termediul selecției grupurilor la care aceștia aparțin. În primul stadiu al eșantionării se alege un eșantion de grupuri de gradul I; apoi din fiecare grup aici ales, se selectează un eșantion de grupuri de gradul II etc. pînă cînd se ajunge la ultimul nivel reprezentat de indivizi. În acest stadiu, se poate include în eșantion întreg grupul imediat superior nivelului individual, ceea ce se întîmplă, de regulă, cînd grupurile sînt de dimensiune redusă.

Un exemplu simplu ne va lămuri și de această dată mult mai bine asupra procedurii decît o expunere abstractă. Să zicem că avem nevoie de un eșantion de studenți la nivel de țară. Cîteva paliere de grupare a studenților pot fi următoarele:

I. Centre universitare

II. Institute de învățămînt superior

III. Facultăți

IV. Grupe de studenți

Am lăsat, după cum se vede, la o parte secțiile și anii de studiu, care ar fi putut mări încă numărul nivelelor de grupare.

Eșantionarea multistadială constă, în cazul de față, în alegerea unui eșantion aleator de centre universitare; în cadrul centrelor alese se vor selecționa institute de învățămînt superior, din care se aleg facultăți, apoi în acestea grupe și, în fine, în cadrul grupei un lot de studenți. Cum spuneam, ne putem opri la penultimul stadiu, adică să includem în eșantion grupa de studenți în întregime. Acest lucru este în deplină concordanță cu rațiunea principală ce ghidează folosirea unei asemenea selecții: efectuarea studiului asupra unei populații cît mai grupate, în scopul economisirii timpului și a mijloacelor materiale. Este clar, în cazul de mai sus, că anchetatorii nu vor mai trebui să se deplaseze în toate centrele universitare, ci doar în cîteva, apoi și aici vor merge doar în unele institute și facultăți. Si în alte cazuri de anchete se pot evidenția avantajele practice ale acestei eșantionări. De pildă, dacă trebuie ales un eșantion dintr-un oraș mare, pe schița localității se pot trasa carouri de o anumită dimensiune, dintre care cîteva se aleg în mod aleator și doar în interiorul acestora se va selecta eșantionul de indivizi.



Operațiile practice legate de eșantionarea multistadială sînt ceva mai complicate decît au lăsat să se înțeleagă rîndurile de mai sus. În primul rînd se pune problema inegalității mărimii grupurilor; ea se rezolvă dînd șanse diferite fiecărui grup de a intra în eșantion, șanse proporționale cu volumul său. Apoi, trebuie efectuate calcule ceva mai complexe legate de reprezentativitatea eșantionului. Aici trebuie reținut că, la mărimi egale, eșantionul multistadial este mai puțin reprezentativ decît cel simplu aleator. Motivul este evident: alegînd grupuri, care sînt în număr mult mai mic decît cel al indivizilor, cresc șansele de a se omite unele categorii din populație.

D. Eșantionarea multifazică. Mărimea eșantionului depinde, printre altele, așa cum spuneam, de caracteristicile populației studiate și de gradul de reprezentativitate urmărit. Dar reprezentativitatea trebuie, la urma urmei, calculată pentru fiecare caracteristică în parte, ea fiind invers proporțională cu dispersia variabilei în cauză. De aici rezultă că dacă se alege eșantionul de așa manieră încît el să aibă o reprezentativitate satisfăcătoare în raport cu variabila cea mai dispersată, atunci pentru celelalte caracteristici mărimea eșantionului devine excesivă, plusul de precizie realizat neavînd o valoare deosebită pentru investigație. Pe de altă parte, în cadrul oricărei cercetări există probleme mai importante - sau cărora sociologul le atasează o importanță mai mare - și altele mai puțin importante. Pentru primele se va urmări, de regulă, obținerea unor estimări foarte precise, în timp ce ultimele sînt abordate satisfăcător și dacă eșantionul este mai mic și nu se realizează o evaluare foarte riguroasă.

În al treilea rînd, tot plecînd de la distincția anterioară între aspectele cercetate pe criteriul importanței, în cazul anchetelor, mai ales, se urmărește nu o dată ca unele chestiuni să beneficieze de un studiu mai intens; dacă, de exemplu, metoda de bază de culegere a datelor este ancheta exată pe un chestionar cu ovasitotalitatea întrebărilor închise și cu formulări standardizate, se simte nevoia ca aspectele de bază să fie abordate și prin alte metode, cum ar fi interviul, care să aducă un plus de cunoaștere pe tărîmul motivațiilor, al atitudinilor etc.



Dar o asemenea investigație intensivă nu este posibilă pentru întreaga masă de indivizi din eșantion căci nu se dispune nici de timp, nici de personal calificat pentru interviuri, nici de posibilități de analiză și prelucrare a unei asemenea cantități de informație nestandardizată-

În toate problemele semnalate mai sus duc la adoptarea unei tactici de cercetare specifice, surprinsă în titlul de eșantionare multifazică. E vorba, în linii mari, de faptul că se alege inițial un eșantion de dimensiune mare, în cadrul căruia se realizează diferite faze ale cercetării; apoi din aceasta se selecționează un subeșantion pentru realizarea altor faze, eșantion ce, la rândul lui, poate fi micșorat printr-o selecție aleatoare pentru alte faze etc. De exemplu pe eșantionul mare putem aplica un chestionar simplu, cu puține întrebări și răspunsuri standardizate, în scopul de a decela unele structuri de bază ale populației. În faza a doua, pe un subeșantion, putem aplica un chestionar mai bogat, cu întrebări deschise, de opinie și atitudine; într-o fază ulterioară, pe un eșantion de mărime și mai redusă se poate aprofunda în continuare cercetarea prin interviuri, bugete de timp, bugete de familie, observație etc.

Aceste faze, logic delimitate aici, se pot realiza simultan în timp. Adică, dacă e vorba de anchetă, operatorul primește instrucțiuni exacte la care dintre indivizii anchetați să aplice numai instrumentele prevăzute pentru prima și a doua ș.a.m.d. Instrumentele, la rândul lor, pot fi separat întocmite sau pe unul și același formular vor fi marcate întrebările specifice diferitelor faze ale cercetării.

Desigur că, din punctul de vedere al preciziei eșantionării în fazele ulterioare, este preferabil ca lucrurile să se deruleze succesiv în timp și nu simultan. Astfel, după culegerea datelor în prima fază și după o prelucrare sumară a lor, vom avea posibilitatea să obținem, pentru următoarele faze, eșantioane prin stratificare, folosind informația mai înainte dobândită. În felul acesta, reprezentativitatea eșantioanelor mai mici din fazele ce urmează va crește sensibil. O asemenea strategie a cercetării îi mărește evident costul și timpul de realizare și deci considerentele practice pot preleva asupra celor de



natură cognitivă.

O precizare importantă se impune și aici. Avînd de a face cu o eșantionare probabilistică, în situația cînd toate fazele sau mai multe faze se realizează simultan, este nevoie ca anchetatorii să primească instrucțiuni foarte exacte legate de identitatea persoanelor ce formează eșantioanele din fazele a doua, a treia etc. Adică, nu trebuie lăsată la latitudinea operatorilor alegerea subiecților chestionați în aceste faze, deoarece astfel se iese din cadrul procedurilor de selecție probabilistă.

În încheierea acestui subparagraf, doresc să subliniez ideea, ce s-a putut deja întrezări pe parcursul său, că cele patru tipuri de eșantionare probabilistă nu se exclud unele pe altele și deci că, în cadrul uneia și aceleiași cercetări, putem avea o împletire a lor. Practic, orice combinație este posibilă: eșantion stratificat bistadial, eșantion stratificat bifazic etc. În astfel de cazuri, una dintre problemele principale este cea a calculului reprezentativității, fiecare procedură făcînd apel la alte formule. Spun aceasta pentru că nu o dată se folosesc astfel de procedee combinate de eșantionare, în timp ce calculele de reprezentativitate sînt făcute ca și cînd eșantioanele ar fi aleatoare elementare.

## 2.2. Eșantionarea nealeatoare

Am arătat mai înainte că principiile de bază ale eșantionării aleatoare sînt: a) posibilitatea calculării șanselor fiecărui individ de a intra în eșantion și b) faptul că aceste șanse nu sînt egale cu zero pentru nici o unitate a populației. Ori de cîte ori o procedură de eșantionare încalcă unul din aceste două - sau ambele - principii ea va trebui considerată nealeatoare. Practic, asemenea proceduri pot căpăta forme de realizare concretă nelimitate și este foarte greu să se facă o tipologizare a lor. Vom avea eșantioane neprobabilitate atunci cînd ieșim în stradă să chestionăm un număr de subiecți, cînd obținem răspunsuri la un chestionar publicat într-o revistă sau un ziar, dacă anchetăm spectatorii prezenți într-un cinematograf sau teatru, dacă ne ducem într-o școală spre a aplica un chestionar



elevilor, cînd, dintr-un județ, alegem 2-3 localități rurale "tipice" pentru investigarea anumitor probleme, dacă ... și exemplele ar putea continua aproape la nesfîrșit, atît de multe variante este posibil să imaginăm sau să le găsim efectiv puse în practică.

Grija principală a oricărui cercetător serios care utilizează în investigație un eșantion nealeator va fi aceea de a limita pe cît posibil subiectivitatea în alegerea unităților ce formează eșantionul. Ingrădirea subiectivității de alegere se poate face prin stabilirea inițială a unor caracteristici pe care să le posede fiecare individ ales. S-a ajuns astfel să se prefigureze și în cadrul selecției nealeatoare o procedură, ale cărei rezultate practice au impus-o, mai cu seamă în cazul anchetelor de opinie, procedură numită eșantionarea pe cote, care va face obiectul rîndurilor ce urmează.

Condiția de bază ce se cere îndeplinită pentru a putea realiza o eșantionare pe cote este cunoașterea prealabilă a cîtorva caracteristici de bază ale populației. În fond, eșantionarea pe cote poate fi considerată ca o eșantionare stratificată neprobabilistă. Cunoșcînd distribuția populației în clasele caracteristicilor de stratificare se va calcula, din mărimea totală a eșantionului, cîți indivizi trebuie aleși din fiecare clasă a acestor caracteristici structurale. Dacă e vorba de o anchetă cu operatori de interviu, cum se întîmplă cel mai adesea, atunci fiecărui operator i se va indica, pe lîngă numărul total de persoane de anchetat, și însușirile pe care acestea trebuie să le posedă, precum și modalitatea concretă de găsire și abordare a lor. Sarcinile ce revin unui singur operator se numesc cote, de unde și numele procedeului.

Să luăm și aici un exemplu, pentru a facilita înțelegerea tehnicii. Să zicem că ne interesează obținerea unui eșantion de 2.000 de studenți dintr-un centru universitar oarecare, spre a fi anchetați pe baza unui chestionar. Să mai presupunem că avem la dispoziție 100 operatori și deci că fiecare dintre aceștia este chemat să realizeze 20 de chestionare. Despre populația studențească a centrului universitar cunoaștem distribuția în funcție de:



- apartenența la cele patru institute de învățămînt superior existente în oraș (notate cu A,B,C și D):

A - 40% ; B - 30% ; C - 20% ; D - 10%

- sex:

băieți - 55% ; fete 45%

- ani de studii:

I - 30% ; II - 25% ; III - 20% ; IV - 15% ; V - 10%

Prin urmare, esantionul nostru de 2000 de studenți va trebui să conțină elemente astfel repartizate:

- după institute:

A - 800

B - 600

C - 400

D - 200

- după sex:

băieți - 1.100

fete - 900

- după ani de studiu:

I - 600

II - 500

III - 400

IV - 300

V - 200

Presupunînd că sarcinile operatorilor sînt uniforme (adică identice pentru toți), fiecare va trebui să chestioneze 20 de studenți, repartizați după cum urmează:

- după institute:

A - 8; B - 6; C - 4; D - 2

- după sex:

băieți: - 11; fete - 9

- după ani de studiu:

I - 6; II - 5; III - 4; IV - 3; V - 2;

Acetoea constituie deci cotele ce revin fiecăruia, urmînd ca în investigarea efectivă anchetatorul să-și aleagă subiecții pe care-i dorește, cu condiția ca în final să fie îndeplini-



nite condițiile de mai sus. Am ales aici, în scop didactic, un caz foarte simplu, dar cred că se înțelege că, în realitate, este inefficient să se repartizeze operatorilor sarcini identice, deoarece aceasta implică obligativitatea ca fiecare operator să meargă în toate cele patru institute. Sarcinile se pot concentra, în cazul de față, prin repartizarea a 40 de operatori exclusiv în institutul A, a 30 de operatori, în B, a 20 în C și a 10 în D.

Criteriile de sarcini de genul celor mai sus ilustrate se numesc independente, deoarece cei trei factori pe baza cărora se constituie cotele nu sînt legați unul de celălalt. Astfel, cei 4 studenți din anul III pot fi aleși tiți dintr-un singur institut sau cîte unul din fiecare; ei pot fi toți de același sex sau 2 băieți și 2 fete etc. În cazul cotelor independente nu se face deci nici o precizare prealabilă asupra raporturilor dintre cote. Desigur că de cele mai multe ori acest lucru poatea avea consecințe neplăcute, în sensul că unele categorii vor apare supra-reprezentate iar altele subreprezentate. De exemplu, în cazul nostru, dacă toți operatorii vor raționa astfel "am de ales 4 studenți din anul III și tot 4 din institutul C, atunci cel mai comod este ca mergînd în institutul C să caut direct anul III și să-mi epuizez de aici sarcina", va rezulta în final că toate opiniile studenților de anul III vor fi caracteristice doar institutului C. Normal că am exagerat puțin aici și că nu toți vor proceda astfel, dar sînt șanse foarte mari ca unii să o facă și atunci răspunsurile în eșantion vor fi distorsionate în aceste alegeri.

Pentru a evita asemenea situații, este preferabil ca să se lucreze cu cote legate, adică să se constituie, de fapt, un tabel de asociere simultană între toate caracteristicile avute în vedere. De pildă, un anchetator poate primi sarcina astfel:

=====									
I N S T I T U T									
A		B		C		D			
An	Sex	Băiat	Fată	Băiat	Fată	Băiat	Fată	Băiat	Fată
I		1	2	1	1	-	1	-	-
II		1	1	1	-	1	-	1	-
III		1	-	1	-	-	1	-	1
IV		1	1	-	-	1	-	-	-
V		-	-	1	1	-	-	-	-
=====									



iar ceilalți în mod similar, în așa fel încît pe ansamblu să ia-  
să atît numărul dorit, dar și anumite corelații între factorii  
de stratificare.

Este clar că nu putem utiliza criteriul corelate decît  
dacă dispunem de o bază statistică după care să putem proiecta  
respectivetele cote, adică atunci cînd știm care sînt asocierile  
efective între caracteristici în populație. Chiar și în cazul în  
care astfel de corelații sînt imprecise, se recomandă totuși o  
formulare cît mai specifică a cotelor, pentru ca operatorul să  
aibă de a face cu cît mai multe elemente constrîngătoare în ale-  
geri, să i se micșoreze pe cît posibil facilitatea de a înlocui  
un element cu altul. Desigur că nu trebuie nici exagerat pe  
această cale; dacă operatorul întîmpină prea multe obstacole în  
depistarea indivizilor cu caracterele indicate, el va avea ten-  
dința de a "fabrica" el însuși asemenea personaje.

De asemenea, se mai recomandă, în achetele de opinie  
bazate pe un eșantion pe cote, ca sarcinile operatorului să fie  
cît mai restrînse (numeric) cu puțină, nedepășind 15-20 de in-  
terviuri. Mai multe motive pledează pentru utilizarea unui număr  
mare de anchetatori cu sarcini reduse fiecare. Mai întîi, alege-  
rile subiective, care nu pot fi totuși înlăturate, se vor orien-  
ta spre un număr mai mare de direcții, care, cu o probabilitate  
ridică, vor duce la compensări reciproce. O sarcină mică re-  
duce, de asemenea, ponderea pe întreg eșantionul a orientărilor  
subiective prea pronunțate ale unor operatori. Nu este tot una  
dacă un asemenea operator realizează 15 sau 150 de chestionare;  
efectul distorsionant este mult mai mare în cazul al doilea. Da-  
că au de făcut un număr prea mare de anchete, operatorii vor  
avea tendința de a le realiza cît mai operativ și asta are efec-  
te negative nu numai în modul de completare a chestionarului, ci  
și în ceea ce privește grija de a selecta persoanele în eșan-  
tion. Cum spuneam, este important să se specifice anchetatorilor  
și modul în care vor lua contact cu persoanele anchetate: acasă,  
la locul de muncă, pe stradă etc., în scopul de a evita contac-  
tarea prea numeroasă a unor tipuri specifice din populație. Or,  
este evident că un operator încărcat cu prea multe sarcini va  
alege, de preferință, cele mai accesibile mijloace de contacta-  
re a subiecților.



În literatura de specialitate, există multe discuții referitoare la avantajele și dezavantajele eșantionării pe cote (în general, a selecției nealeatoare), în raport cu cea aleatoare. La prima vedere, faptul că pentru aceasta din urmă este posibilă calcularea reprezentativității iar pentru prima nu, poate părea un avantaj decisiv, care să încline complet balanța în favoarea studiilor pe eșantioane probabiliste. Iar dacă argumentul nu pare totuși convingător se mai pot aduce și altele în sprijin. E vorba, mai întâi de dificultățile de a construi cote în raport cu cele mai relevante caracteristici pentru studiu, dat fiind că nu se cunoaște, de cele mai multe ori, distribuția lor în populație. În felul acesta este greu să se îngrădească subiectivitatea în alegere, subiectivitate care oricum rămâne în orice selecție nealeatoare. În fine, un alt argument de marcă împotriva acestui gen de selecție este dat de imposibilitatea realizării unui control eficient a muncii de teren a operatorilor. În eșantionul aleator, cel care conduce cercetarea cunoaște identitatea fiecărei persoane ce trebuie anchetată și deci poate merge pe urma operatorilor spre a verifica dacă și cum s-au realizat anchetele. În eșantionarea pe cote, mai cu seamă dacă chestionarul este anonim, pierdem orice posibilitate de control, nefiind siguri nici măcar dacă persoană chestionată există efectiv.

Cu toate acestea, studiile pe baza eșantioanelor nealeatoare nu numai că nu sînt rare, dar probabil, dacă s-ar face o statistică, s-ar constata că ele sînt întîlnite cu preponderență în sociologie și psihologia socială. Lucrul se întîmplă, de regulă, prin alegerea deliberată a acestui gen de selecție, dar mai există și cazuri cînd cercetătorii au numai impresia sau lasă numai impresia că au ales un eșantion aleator, cînd de fapt, situația este cu totul alta.

Lăsînd la o parte acest din urmă caz, să vedem care sînt motivele principale ce duc totuși la o largă utilizare conștientă a eșantioanelor nealeatoare, pe cote, în speță. Trebuie să înțelegem, în prealabil, că factorul principal de care depinde modalitatea de efectuare a unei cercetări este eficiența pe care dorim s-o aibă această investigație. Noțiunea de eficiență înglobează, în primul rînd, rezultatele teoretice sau aplicative ale cercetării, rezultate ce depind în bună măsură de acu-



teșea eșantionării în studiile selective. Dar alte laturi, care pot fi neglijate, sînt costul cercetării și promptitudinea (rapiditatea) cu care se ajunge la rezultatele finale. Desigur că ponderea fiecărui factor depinde de genul de cercetare efectuată, de scopul ei etc. În cazul unor cercetări cu caracter aplicativ, comandate de un beneficiar, ultimii doi factori acționează cu o forță deosebită.

Am făcut această paranteză pentru a înțelege mai bine de ce nu sînt rare situațiile cînd se preferă eșantioane pe cote. Într-adevăr, costul realizării acestora este mult mai mic, preparativele sînt relativ sumare și deci promptitudinea cercetării este deosebită. Faptul e ușor de ilustrat prin investigațiile efectuate de institutele occidentale de sondare a opiniei publice asupra unor probleme precum: audiența emisiunilor radio și TV, preferințele în campaniile electorale etc. Este clar că asemenea informații trebuie culese și prelucrate rapid, altfel datele își pierd orice valoare. Or, numai un eșantion pe cote este capabil să furnizeze informația dorită în asemenea cazuri. Cu siguranță că și eșantionarea pe cote presupune o fază de pregătire atentă, bazată pe cunoașterea realității sociale investigate, fapt confirmat și de secretul în care este învăluită această etapă a investigației, în cazul institutelor de opinie publică occidentale. Într-adevăr, procedura de alegere a unui eșantion puternic reprezentativ la nivel național pare să fie operația "tehnologică" din științele sociale al cărei secret este cel mai bine păzit. Experiența de pînă acum o dovedește că eșantionarea pe cote este minuțios preparată, datele obținute se pot extrapola la populația totală cu o marjă de eroare similară celei calculate pentru eșantioane probabiliste de aceeași talie. Astfel, ca să reiau exemplul institutelor menționate, în țări de mărimea Angliei, Franței sau R.F. a Germaniei, cu un eșantion național mai mic de 2000 de persoane, studiat în preajma alegerilor, se pronostichează rezultatele cu o precizie ce merge uneori pînă la mai puțin de 1%.

Asemenea performanțe sînt deseori subliniate de practicieni, în ciuda criticilor statisticienilor, pentru care singura selecția aleatoare rămîne capabilă să dea rezultate extrapolabile. Tot practicienii mai remarcă, în mod judicios așa spune, că



de fapt eşantionul probabilistic ales nu este niciodată în întregime investigat (cînd e vorba de anchete, evident). Într-adevăr există mereu un procent, ce deseori depăşeşte 30%, de indivizi care nu pot fi găsiţi acasă, sînt inapţi pentru răspunsuri, refuză să colaboreze etc. Chiar dacă numărul lor poate fi redus prin vizite repetate, este clar că fenomenul nu poate fi în întregime eliminat. E.Noelle (1966) aprecia că valoarea de 80% din eşantion poate fi greu atinsă, chiar dacă se fac pînă la 3 vizite. Apoi insistenţele pe lîngă persoanele care nu doresc să răspundă au mai degrabă un efect negativ asupra rezultatelor cercetării, iar cele pe lîngă persoanele inapte sînt cel puţin deplasate. Prin urmare, eşantionul, oricît de corect ar fi iniţial, şi mai ales dacă a fost corect ales, rămîne pînă la urmă trunchiat, fiind astfel purtătorul unor posibile erori sistematice,

Sînt, de asemenea, înclinat să dau dreptate celor care susţin că, de fapt, în anchetele pe bază de eşantion, erorile cele mai importante sînt departe de a fi cele datorate eşantionării, ele provenind mai degrabă din modul concret de administrare a chestionarului, din munca operatorilor, din felul în care se dau şi se înregistrează răspunsurile.

În fine, dar nu în ultimul rînd ca importanţă, ci poate chiar în primul, eşantionarea nonprobabilistică se impune din cauza inexistenţei unui cadru de eşantionare suficient de bun sau adus la zi şi a eforturilor enorme pe care un sociolog ar trebui să le facă pentru a-şi construi singur un astfel de cadru. Listele cu populaţia sînt de cele mai multe ori vechi şi incomplete, iar unele satisfăcătoare existente pot să nu-i fie accesibile sociologului.

Iată cîteva motive care explică şi pledează pentru folosirea pe scară largă a eşantioanelor nealeatoare. Expunerea lor şi sublinierea unor avantaje deloc neglijabile ale acestui gen de selecţie nu înseamnă o respingere de principiu a metodelor aleatoare. Dimpotrivă, gîndul oricărui sociolog trebuie îndreptat spre exploatarea tuturor posibilităţilor de a întocmi un eşantion probabilist. Numai dacă obstacole ce stau în calea acestui plan sînt insurmontabile, în timpul şi cu mijloacele avute la dispoziţie, trebuie să apeleze, cu prudenţă şi vigilenţă, la folosirea unuia aleator.



### 2.3. Eșantioane fixe (panel)

Ieșind din firul logic ce a ghidat clasificarea folosită pînă aici (eșantionare aleatoare și nealeatoare), voi pune, foarte pe scurt în evidență un tip deosebit de eșantioane, care nu se disting prin modul în care au fost construite, ci prin ce de utilizare a lor. Eșantioanele fixe (sau panel, după un termen englezesc) sînt acelea care, indiferent de modalitatea lor de alegere, servesc unor studii repetate la diferite intervale de timp.

În cazul în care ele sînt constituite din indivizi umani, ele ridică unele probleme specifice, ce se cer menționate aici. Mai întîi, este limpede că repetarea unor cercetări, identice sau diferite ca problematică, asupra aceluiași eșantion aduce avantaje în ceea ce privește reducerea costului cercetării a timpului și, dacă reprezentativitatea eșantionului s-a dovedit corespunzătoare, și în privința acurateții rezultatelor. Odată ce indivizii au fost supuși unei investigații, nu numai că nu mai trebuie din nou selecționați - cu toate riscurile și greutățile caracteristice acestei operațiuni - dar și o serie de probleme legate de contactul cercetător-subiect angajarea subiectului într-o discuție deschisă etc. dispar.

Probabil însă că avantajul cel mai semnificativ al eșantioanelor fixe se manifestă atunci cînd se repetă o anchetă, în scopul evaluării și explicării schimbărilor sociale petrecute în intervalul de timp dintre cele două momente. Desigur, că și studiul pe două eșantioane diferite evidențiază schimbările, dar atunci cînd eșantionul rămîne același se pot pune în evidență indivizii care și-au modificat "comportamentul" în răstimpul dat, în raport cu o caracteristică studiată, și din analiza situațiilor lor se vor putea trage concluzii referitoare la factorii de schimbare.

Nu trebuie totuși supraestimate virtuțile eșantioanelor panel. Multe din avantajele lor se transformă treptat în dezavantaje. Oamenii se pot plictisi de anchete repetate și vor începe să refuze de a mai răspunde; unii mor, alții se mută, alții devin indisponibili pentru anchetă. Astfel, încet, eșantionul în



cepe treptat să se degradeze "fizic" și trebuie mereu reîmpros-  
pătat, ceea ce evident îi anulează o parte din avantaje.

Pe lângă degradarea "fizică", se poate vorbi și de una  
"morală", în sensul că oamenii fie că vor deveni plătisiți și  
vor începe să răspundă tot mai superficial la întrebări, fie că  
vor suferi un fenomen de "condiționare". Știind că va fi inter-  
ogat despre anume probleme, omul, de cele mai multe ori, își  
schimbă comportamentul față de factorii respectivi; și nu numai  
comportamentul, dar și atitudinile, opiniile, convingerile etc.  
De exemplu, știind că va fi interviuată periodic în legătură  
cu programele de televiziune, orice persoană va avea tendința să  
urmărească mai des decât în mod normal aceste emisiuni, să le  
județe cu ochiul celui care va trebui să le aprecieze calitatea  
în fața cuiva etc. și astfel este foarte probabil că organizato-  
rul sondajului va constata rapid că în eșantionul său crește  
timpul de vizionare a programelor și se schimbă aprecierile a-  
cestora, când în realitate (adică în populație) lucrurile pot fi  
cu totul altele.

În ciuda celor de mai sus, eșantionarea panel rămâne  
o modalitate foarte eficientă de culegere rapidă a unor informa-  
ții consistente, mai ales atunci când este vorba de a analiza  
factorii care influențează schimbările în opiniile, atitudinile  
și comportamentul indivizilor umani în raport cu anumite fenome-  
ne sociale care se derulează cu viteză ridicată și nu este timp  
pentru punerea în aplicare a unor strategii de cercetare mai com-  
plexe, dar mai greoaie, mai puțin prompte în aplicare.



## Cap. XII. ASPECTELE MATEMATICE ALE ESANTIONARII

### 1. Determinarea mărimii și reprezentativității esan- tsonului. Estimarea mediilor și proporțiilor.

În general, scopul studiului pe bază de eșantion îl constituie estimarea unor mărimi ale populației, de regulă, medii sau proporții. Spun estimare și nu determinare, datorită faptului că prin intermediul unui eșantion nu putem obține valoarea exactă urmărită - dacă în general are sens să vorbim de o valoare exactă - decât în cazuri cu totul excepționale și datorită unui hazard pe care nu ne putem prea mult baza. Dar pentru nevoile practicii este mai mult decât suficientă obținerea unor valori estimative, apropiate de cea care admitem că există realmente în populație.

Să presupunem că am efectuat un studiu exhaustiv (asupra întregii populații) și am obținut următoarea caracteristică I:

Valori	$x_1$	$x_2$	$x_3$	...	$x_s$	Total
Frecvențe	$k_1$	$k_2$	$k_3$	...	$k_s$	$N$

În acest caz se poate calcula media exactă a variabilei X:

$$\bar{x} = \frac{k_1 x_1 + k_2 x_2 + \dots + k_s x_s}{N}$$

precum și alte mărimi, precum abaterea standard  $\sigma_x$ . Efectuînd o cercetare selectivă, nu vom studia pe toți cei  $N$  indivizi din populație, ci numai o parte a lor, să zicem un număr  $n$ . Pentru aceștia obținem o caracteristică  $X'$  de forma:

Valori	$x_1$	$x_2$	$x_3$	...	$x_s$	Total
Frecvențe	$k'_1$	$k'_2$	$k'_3$	...	$k'_s$	$n$

cea medie va fi:



$$\bar{x}' = \frac{k_1'x_1 + k_2'x_2 + \dots + k_s'x_s}{n}$$

aproape sigur diferită de  $\bar{x}$ .

Diferența  $\bar{x} - \bar{x}'$  se numește eroarea introdusă de selecția care a condus la mărimea  $\bar{x}'$ . Este evident că pentru orice cercetare este preferabil să se obțină o eroare cât mai mică, adică cele două medii să fie cât mai apropiate. Dar tot atât de evident este că dacă eroarea nu depășește o anumită limită, estimarea lui  $\bar{x}$  prin  $\bar{x}'$  este absolut satisfăcătoare pentru scopurile cercetării și depunerea unor eforturi suplimentare pentru a reduce diferența  $\bar{x} - \bar{x}'$  nu este rentabilă, deoarece se măresc cheltuielile și timpul reclamate de investigație.

De pildă, dacă urmărim evaluarea câștigului mediu lunar realizat de o colectivitate de muncitori, în unele cazuri este suficient să știm că eroarea comisă nu depășește 50 lei. Aceasta înseamnă că dacă obținem în eșantion un câștig mediu de 2200 lei, vom ști că cel care caracterizează populația va fi de această mărime  $\pm 50$  lei. În mod identic se pune problema evaluării unei proporții. Dacă ne satisface o eroare de 1%, atunci prin cercetarea selectivă se va urmări respectarea condițiilor ca proporția obținută să nu difere de cea "adevărată" cu mai mult de un procent:  $|p - p'| < 0,01$ . De exemplu, dacă în eșantion găsim o proporție de activi de 0,45 (45%), atunci în populație ea va fi de  $0,45 \pm 0,01$ , deci între 44% și 46%.

De fapt, afirmațiile de mai sus trebuie - în aparență puțin dar în realitate fundamental - corijate. Dat fiind că valoarea adevărată din populație (media  $\bar{x}$  sau proporția  $p$ ) este necunoscută (altfel n-ar avea rost studiul), nu se va putea afirma niciodată cu certitudine că diferența  $|\bar{x} - \bar{x}'|$  sau  $|p - p'|$  este mai mică decât o anumită valoare. Teoria matematică a probabilităților ne ajută la stabilirea unui alt gen de afirmație, și anume la a putea spune că respectiva diferență este mai mică decât o mărime  $\Delta$  aleasă după exigențele de precizie ale cercetării, numită eroare maximă admisă, cu o anumită probabilitate  $P_0$ , suficient de ridicată pentru a o putea considera ca o cvasicertitudine. Mai scriem această probabilitate, numită și nivel de încredere, sub forma  $P(|\bar{x} - \bar{x}'| < \Delta)$  și o citim: "probabilitatea ca me-



dă din populație și cea din eșantion să fie mai mică decât  $\Delta$ . Nivelul de încredere cel mai jos utilizat este de 95% (sau 0,95) iar în practică, pentru cazuri de exigență sporită, se mai folosesc cele de 99% (0,99) și 99,9% (0,999).

Se vede deci că precizia oricărei eșantionări privește două elemente: eroarea  $\Delta$  și probabilitatea  $P$ . Ambele elemente de incertitudine,  $\Delta$  și  $P$ , se datorează faptului că eșantionul este doar o parte a populației, al doilea element fiind, în același timp, determinat și de necunoașterea mărimii pe care vrem s-o estimăm. Cuplul  $(\Delta, P)$  constituie ceea ce, de obicei, se numește reprezentativitatea eșantionului. Deci reprezentativitatea privește în același timp intervalul în jurul mediei din eșantion  $(\bar{x} - \Delta, \bar{x} + \Delta)$  - numit și interval de confidență - cât și probabilitatea cu care putem afirma că media din populație,  $\bar{x}$ , cade în acest interval.

Vom spune că un eșantion  $A$  este mai reprezentativ decât unul  $B$  dacă, la același interval de confidență, nivelul de probabilitate al lui  $A$  este mai ridicat decât cel al lui  $B$  sau dacă, la nivele egale de încredere, intervalul de confidență al lui  $A$  este mai mic sau dacă, simultan, nivelul de încredere e mai mare și intervalul mai mic. Un eșantion poate fi numit pur și simplu reprezentativ (decă fără comparație cu altul) dacă cele două mărimi se încadrează în limitele cerute de necesitățile cercetării, adică  $\Delta$  este mai mic decât eroarea maximă admisă iar  $P$  mai mare decât gradul minim de probabilitate admis: 0,95.

Este clar, din afirmația făcută puțin mai sus asupra surselor incertitudinii, că mărirea gradului de reprezentativitate al eșantionului se poate realiza, în condiții egale, prin sporirea volumului eșantionului. Atunci când  $n$  se apropie de  $N$ ,  $\Delta$  va tinde spre 0 (exactitudine) iar  $P$  spre 1 (certitudine). Deci este o funcție descrescătoare de  $n$  iar  $P$  una crescătoare. Dar la un  $n$  fixat mărimile  $\Delta$  și  $P$  nu sînt independente una de alta. Reducînd eroarea  $\Delta$  (decă sporind exactitatea estimării) se micșorează probabilitatea  $P$  (adică șansele ca această exactitate să se realizeze) și invers, mărind probabilitatea, mărim și eroarea. Este, dacă vreți, o analogie, evident formală, cu celebrul principiu al incertitudinii al lui Heisenberg.



În ceea ce privește reprezentativitatea eșantionului, deci mărimile  $\Delta$  și  $P$ , se pot pune două probleme:

1) Să se determine mărimea  $n$  a eșantionului astfel încât să se realizeze reprezentativitatea aleasă (un  $\Delta$  și un  $P$ ). De exemplu, dacă reluăm cazul ilustrat mai înainte, să se determine numărul muncitorilor ce trebuie anchetați, pentru a putea spune cu o probabilitate de, să zicem, 95% ( $P = 0,95$ ) că media din populație nu diferă cu mai mult de 50 lei de cea din eșantion ( $\Delta = 50$  lei).

2) Având dat un eșantion de mărime  $n$ , să se aprecieze reprezentativitatea acestuia, deci să se calculeze eroarea maximă la un anumit nivel de probabilitate. Adică, dacă am investigat 1000 de muncitori și am obținut o medie de venituri de 2200 lei, să se vadă la nivelul de probabilitate de 0,95 (sau altul mai mare) în ce interval de confidență cade media din populație.

Prima problemă se ridică în faza de proiectare a cercetării iar a doua în cea de prelucrare a datelor. Așa cu, se va vedea mai jos, este greșit să se creadă că rezolvarea primeia implică și rezolvarea secundeia, calculele trebuind făcute și înainte și după efectuarea investigației concrete. De asemenea, trebuie subliniat că anumite formulări, din păcate destul de des întâlnite în studiile sociologice de teren, de genul "s-a studiat problema cutare pe un eșantion reprezentativ" sînt absolut fără nici o semnificație dacă nu se precizează cei doi indicatori ai reprezentativității:  $\Delta$  și  $P$ .

Problemele propuse se rezolvă în felul următor. Să ne imaginăm că am studiat un eșantion A de volum  $n$  și am obținut o medie  $\bar{x}'_a$ . Extragem apoi un alt eșantion B, de același volum, și obținem o medie  $\bar{x}'_b$ . Repetînd indefinit operația, se va obține un șir de valori:  $\bar{x}'_a, \bar{x}'_b, \bar{x}'_c, \dots$ , care, fiecare, constituie o estimatie a valorii din populație  $\bar{x}$ . Se ajunge astfel la o variabilă statistică cu valorile de mai sus, care, la rîndul său, va avea o medie (media tuturor mediilor eșantioanelor), notată cu  $\bar{x}'$  și o abatere standard  $\sigma_{\bar{x}'}$ . Primul indicator arată în jurul cărei valori se grupează mediile diferitelor eșantioane iar ultimul, gradul de împrăștiere față de poziția centrală.

Teoria matematică a probabilităților demonstrează,



pentru variabila construită mai sus, următoarele trei propoziții fundamentale, necesare și suficiente pentru rezolvarea problemelor enunțate:

a) Valoarea medie a distribuției mediilor eșantioanelor este egală cu media populației:

$$\bar{\bar{x}}' = \bar{x} \quad (1)$$

Adică luând valorile medii obținute în fiecare eșantion în parte și făcând media acestora, se ajunge exact la valoarea căutată din populație.

b) Varianta (pătratul abaterii standard) distribuției mediilor eșantioanelor este de n ori mai mică decât varianta distribuției indivizilor în populație:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \quad (2)$$

Se observă deci că împrăștierea mediilor eșantioanelor în jurul lui  $\bar{x}$  este relativ mică și depinde de dispersia variabilei studiate (directe proporțional) și de volumul eșantionului (invers proporțional). Este prin urmare evident că mărirea numărului de indivizi în eșantion conduce la reducerea împrăștierei, deci la o "strângere" a mediilor eșantioanelor în jurul mediei populației.

Formula de mai sus este, strict vorbind, valabilă numai în cazul selecției simple aleatoare repetate. Pentru selecție nerepetată, intervine un factor de corecție ce modifică formula astfel:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \cdot \frac{N - n}{N - 1}$$

unde  $N$  este volumul populației. Pentru a nu complica prea mult expunerea, în cele ce urmează se va folosi doar formula mai simplă (2).

c) Variabila statistică  $\bar{x}'$ , adică mediile eșantioanelor, se distribuie după o curbă normală.



Cîteva cuvinte explicative se impun în legătură cu curba normală. Este vorba de o funcție matematică dată de formula:

$$f(u) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{u - \bar{u}}{\sigma} \right)^2}$$

unde  $u$  este variabila independentă,  $\bar{u}$  - valoarea medie, iar  $\sigma$  abaterea standard;  $e$  fiind cunoscuta constantă matematică, baza logaritmilor naturali. (Nu am notat cu  $x$ , cum este obiceiul, variabila independentă pentru a nu o confunda cu variabila urmărită în cercetarea noastră, despre care nu putem presupune că ar fi normală). Graficul funcției  $f(u)$  este cunoscuta curbă a lui Gauss, sub formă de clopot, simetrică în raport cu valoarea medie și tinzînd asimptotic la zero, spre  $-\infty$  și  $+\infty$ . Forma mai plată sau mai ascuțită a curbei depinde de celălalt parametru care apare în formulă: abaterea standard. Cu cît ea este mai mare, cîtît curba este mai plată și se apropie mai greu de axa absciselor (fig.1).

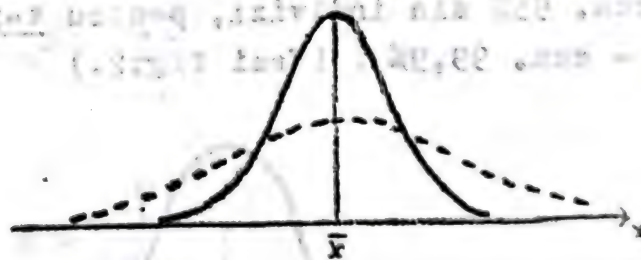


Fig.1. Două curbe normale, cu aceeași medie și abateri standard diferite.

Ca orice funcție de distribuție, și aceasta are proprietatea că aria determinată de ea și axa absciselor, între două puncte oarecare  $u_1$  și  $u_2$  adică integrala funcției de la  $u_1$  la  $u_2$  dă probabilitatea ca un individ statistic oarecare să ia o valoare cuprinsă între  $u_1$  și  $u_2$ . Deci:



$$P(u_1 \leq u \leq u_2) = \int_{u_1}^{u_2} f(u) du$$

În cazul funcției normale, dacă valorile  $u_1$  și  $u_2$  sînt simetrice în raport cu media, adică formează un interval al cărui punct central este chiar media, se pot găsi probabilitățile respective cu ajutorul unor tabele, în funcție de lungime intervalului respectiv. Aceste probabilități depind de forma curbei, adică de abaterea standard, dar valorile vor deveni independente de respectivul parametru, dacă lungimea intervalului este măsurată totuși în abateri standard. Valorile  $u_1$  și  $u_2$  sînt de forma:

$$u_1 = \bar{u} - t\sigma$$

$$u_2 = \bar{u} + t\sigma$$

unde  $t$  este un număr care arată cu cîte abateri standard capetele intervalului se află la stînga și la dreapta mediei. Deci curbele normale au remarcabila proprietate că toate conțin, într-un interval de acest gen, aceeași proporție de indivizi. Astfel se constată că pentru  $t=2$ , deci pentru un interval de două abateri standard în stînga și dreapta mediei, corespunde o putere de cuprindere de cca. 95% din indivizi, pentru  $t=2,6$  - cca. 99% și pentru  $t=3,3$  - cca. 99,9%. (Vezi fig.2.)

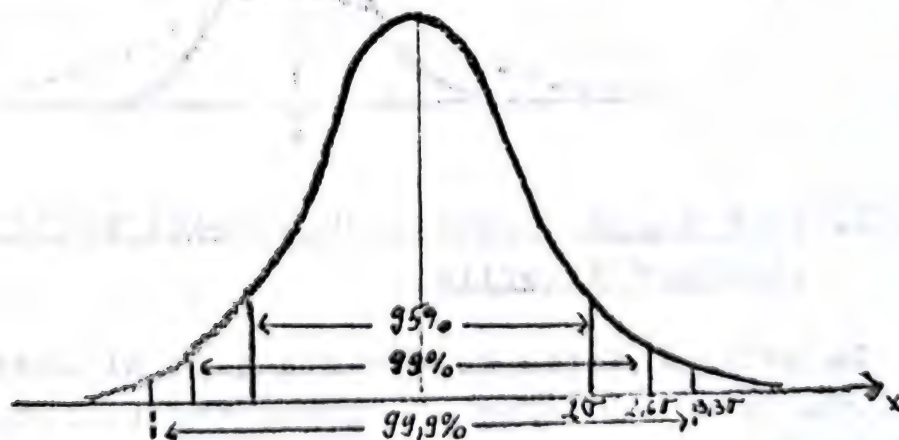


Fig.2. Procentul din aria unei curbe normale



Altfel spus, pentru orice distribuție normală, există 95% șanse ca un individ să se abată de la medie cu mai puțin de două abateri standard, 99% să se abată cu mai puțin de 2,6  $\sigma$  și 99,9% cu mai puțin de 3,3  $\sigma$ .

Să revenim acum la eșantioanele noastre imaginare. Mediile lor urmează deci o curbă normală, în jurul mediei populației, pe care fiecare medie de eșantion o estimează. Noi nu vom ști niciodată exact cât de bună este estimația realizată cu un eșantion anume, cel pe care-l folosim de fapt în cercetare. Stim însă acum câteva lucruri esențiale. Mai întâi, formula (2) ne dă varianța mediilor de eșantioane. Abaterea standard respectivă arată gradul de precizie al eșantionării folosite, reflectând eroarea cea mai probabilă, motiv pentru care poartă denumirea de eroare standard. O vom nota, din motive de comoditate a scrierii cu  $e$ . În al doilea rând, din considerațiile anterioare rezultă că există 95% șanse ca un eșantion oarecare să producă o eroare mai mică decât două erori standard ( $2e$ ), 99% șanse ca eroarea să fie mai mică decât 2,6 $e$  și 99,9% șanse ca ea să nu depășească 3,3 $e$ . (Se observă că valorile alese pentru  $t$ : 2; 2,6 și 3,3 s-au luat pentru a cădea peste nivelele de probabilitate utilizate mai des).

Cele două probleme enunțate se rezolvă acum foarte simplu. Observăm că formula (2) devine, după schimbarea notației:

$$e^2 = \frac{\sum x^2}{n}$$

(3)

sau

$$e = \frac{\sqrt{\sum x^2}}{\sqrt{n}}$$

(4)

#### A. Determinarea mărimii eșantionului

Din formula (3) se scoate  $n$  și se obține:

$$n = \frac{\sum x^2}{e^2}$$

(5)

Mărimea  $e$  se determină în funcție de exigențele cercetării. Se stabilește mai întâi eroarea maximă admisă,  $\Delta$ , care co-



roborată cu un nivel de probabilitate conduce la o valoare pentru  $e$ . În adevăr, dacă alegem  $P = 0,95$ , atunci  $\Delta = 2e$  și deci  $e = \Delta/2$ .

Ceea ce nu cunoaște, însă din formula (5) este valoarea lui  $\sigma_x$ , adică abaterea standard în populație. În etapa inițială a cercetării, această mărime se evaluează, mai exact se înlocuiește cu o valoare despre care să fim siguri că nu va fi depășită în realitate. Așa după cum se vede din formula (5), cu cât vom da lui  $\sigma_x$  o valoare mai mare cu atât volumul eșantionului se va mări și precizia va spori și ea. Deci singura grijă ce trebuie avută e să nu subestimăm pe  $\sigma_x$ .

Exemplu. Urmărim să estimăm cu un eșantion vîrsta medie a unei colectivități de studenți, a căror vîrstă știm că variază aproximativ între 19 și 25 ani. Imprăștierea cea mai mare s-ar obține atunci cînd jumătate dintre indivizi ar fi de 19 ani iar ceilalți de 25 de ani, fiind atunci de 22 ani. S-ar ajunge la o abatere standard de 3 ani (22 - 19 sau 25 - 22), mărime pe care o vom folosi în determinarea lui  $n$ . Drept eroare maximă admisă vom alege 3 luni, adică un sfert de an:

$$\Delta = 0,25$$

La nivelul de probabilitate de 0,95 această mărime conține două erori standard. Deci:

$$e = 0,25 : 2 = 0,125$$

Atunci:

$$n = \frac{3^2}{0,125^2} = \frac{9}{0,0156} = 576$$

cifră care se poate rotunji la 600 pentru mai multă siguranță.

#### B. Determinarea reprezentativității eșantionului

Prin determinarea reprezentativității eșantionului înțelegem, așa cum am spus, calculul erorii maxime admise,  $\Delta$ , și alegem nivelului de probabilitate  $P$ . Acest lucru se face plecînd de la valoarea erorii standard,  $e$ , dată de formula (4). Dar nici acum la sfîrșitul cercetării nu sîntem în posesia valorii exacte  $\sigma_x$ . În schimb, cunoaștem abaterea standard a variabilei  $X'$  în



cadru eșantionului studiat:  $\sqrt{x}$ . Teoria statisticii matematice arată că aceasta reprezintă o estimare suficient de bună a lui  $\sigma_x$ , putînd fi folosită în formulă. Este clar că se ajunge doar la o estimare a erorii standard, însă, pentru simplificarea limbajului, vom considera, în cele ce urmează, că am obținut chiar eroarea standard.

În rezumat: se determină eroarea standard  $e$ , se alege un nivel de probabilitate  $P$ , căruia îi corespunde un număr  $t$ , ce înmulțit cu  $e$ , va conduce la  $\Delta$ , deci la un interval de încredere.

Să reluăm exemplul anterior. Să presupunem că am înregistrat vîrsta celor 600 de studenți anchetați, obținînd o medie:

$$\bar{x} = 22,5 \text{ ani}$$

și o abatere standard:

$$\sigma_x = 1,5 \text{ ani.}$$

Eroarea standard va fi:

$$e = \frac{1,5}{\sqrt{600}} = 0,06$$

Pentru diferite nivele de probabilitate, vom avea:

$P = 0,95$	$\Delta = 2e = 2 \cdot 0,06 = 0,12 \text{ ani}$
$P = 0,99$	$\Delta = 2,6e = 2,6 \cdot 0,06 \approx 0,16 \text{ ani}$
$P = 0,999$	$\Delta = 3,3e = 3,3 \cdot 0,06 \approx 0,2 \text{ ani}$

Prin urmare, sînt 95% șanse ca vîrsta medie a studenților să fie cuprinsă între  $22,5 - 0,12$  ani și  $22,5 + 0,12$  ani, adică în intervalul (22,38, 22,62), 99% în intervalul mai larg (22,34, 22,66) și 99,9% în intervalul și mai mare (22,3, 22,7).

Un alt exemplu. Să se determine volumul eșantionului necesar pentru estimarea mediei notelor obținute de studenți la o sesiune de examene, cu o precizie de 1/10 de notă.

Avem deci  $\Delta = 0,1$ . Dacă lucrăm cu nivelul de probabilitate de 0,95, atunci  $\Delta = 2e$  și, în consecință

$$e = 0,05$$



Notele studenților sînt cuprinse practic între 4 și 10. Abaterea standard maximă poate fi 3, corespunzînd unei împărțiri a studenților în două clase egale: unii de nota 4, alții de 10. Stim însă că o asemenea situație este absurdă, putînd reduce, de la bun început și fără nici o teamă, cifra la 2. Atunci:

$$n = \frac{\sigma_x^2}{e^2} = \frac{2^2}{0,05^2} = \frac{4}{0,0025} = \frac{40000}{25} = 1600$$

Studiînd un eșantion de 1600 studenți, să presupunem că am obținut o medie a notelor:

$$\bar{x}' = 7,80$$

și o abatere standard în eșantion:

$$\sigma_{x'} = 1,6 \text{ note}$$

Cu această din urmă valoare estimăm eroarea standard:

$$e = \frac{1,6}{\sqrt{1600}} = 0,04$$

În situația de față, putem afirma că sînt 95% șanse ca media reală a notelor să fie cuprinsă între  $7,80 - 2 \cdot 0,04$  și  $7,80 + 2 \cdot 0,04$ , adică în intervalul (7,72, 7,88).

### Estimarea proporțiilor

Din punct de vedere formal, problemele de estimare a proporțiilor prin eșantionare se pot reduce la cele similare pentru medii. Într-adevăr, avînd un atribut A, vom acorda valoarea 1 tuturor indivizilor care posedă atributul respectiv și 0 celor care nu-l posedă. Se formează atunci în populație următoarea caracteristică:



Valori	1	0	Total
Frecvențe	$k_1$	$k_2$	$N$

Să notăm:

$$p = \frac{k_1}{N}, \quad q = \frac{k_2}{N}$$

unde  $p$  reprezintă proporția celor care posedă atributul  $A$  și  $q$  a celor care nu-l posedă, cu proprietatea evidentă că  $p + q = 1$ .

Media caracteristicii va fi:

$$\bar{x} = \frac{k_1 \cdot 1 + k_2 \cdot 0}{N} = \frac{k_1}{N} = p$$

Deci media caracteristicii este tocmai  $p$  sau, altfel spus, caracteristica a fost construită de așa manieră încât proporția  $p$  să devină o medie. Se poate calcula și abaterea standard:

$$\begin{aligned} \sigma_x &= \sqrt{\frac{(1-p)^2 k_1 + (0-p)^2 k_2}{N}} = \sqrt{(1-p)^2 \frac{k_1}{N} + p^2 \frac{k_2}{N}} = \\ &= \sqrt{(1-p)^2 p + p^2 q} \end{aligned}$$

Dar fiindcă

$$1 - p = q$$

$$\sigma_x = \sqrt{q^2 p + p^2 q} = \sqrt{pq(q+p)} = \sqrt{pq}$$

Deci

$$\sigma_x = \sqrt{pq} = \sqrt{p(1-p)}$$

Așadar, orice proporție  $p$  poate fi interpretată ca valoare medie a unei variabile cu valorile 1 și 0 și abatere standard dată de formula de mai sus. Fiindcă în problemele de estimare se pune, cum s-a văzut, problema găsirii unei valori maxime a abaterii standard, vom menționa că, pentru cazul de față, se poate demonstra simplu că  $\sigma_x$  atinge valoarea maximă atunci când  $p=0,5$ ,



deci cînd jumătate din indivizi posedă atributul respectiv și jumătate nu-l posedă. Această valoare maximă va fi:

$$\sigma_x = \sqrt{0,5(1-0,5)} = \sqrt{0,5 \cdot 0,5} = 0,5$$

Cu aceste precizări putem trece la prezentarea unui exemplu de estimare a proporțiilor. Să zice, că ne interesează aflarea procentului de indivizi dintr-o populație, care împărtășesc o anumită opinie O. Fie  $p$  proporția în populație și  $p'$  cea obținută în eșantion. Dorim ca diferența dintre  $p$  și  $p'$  să nu depășească 2%, adică, în limbajul proporțiilor, 0,02. Deci:

$$\Delta = 0,02 \quad e = 0,01 \quad \text{pentru } P = 0,95$$

Atunci volumul eșantionului va fi:

$$n = \frac{0,5^2}{0,01^2} = \frac{0,25}{0,0001} = 2500$$

Să presupunem în continuare că ancheta desfășurată pe eșantionul de 2500 indivizi conduce la rezultatul că 20% dintre aceștia împărtășesc opinia respectivă. Deci:

$$p' = 0,2 \quad \text{și} \quad q' = 0,8$$

Abaterrea standard în eșantion va fi:

$$\sigma_{x'} = \sqrt{p'q'} = \sqrt{0,2 \cdot 0,8} = \sqrt{0,16} = 0,4$$

Cu ajutorul ei estimăm eroarea standard:

$$e = \frac{0,4}{2500} = \frac{0,4}{50} = 0,008$$

Pentru  $P = 0,95$ ,  $\Delta = 2e = 2 \cdot 0,008 = 0,016$  sau 1,6%.

În consecință, sînt 95% șanse ca procentul indivizilor din populație care împărtășesc opinia O să fie cuprins între  $20-1,6 = 18,4\%$  și  $20+1,6 = 21,6\%$ . Dacă dorim să ridicăm nivelul de încredere la 99%, atunci eroarea se va ridica și ea la aproximativ 2% ( $2,6 \times 0,008$ ) și deci intervalul de confidență va fi (18%, 22%).



## 2. Teste de semnificație

### 2.1. Semnificația diferenței mediilor sau proporțiilor

De multe ori în lucrările sociologice întâlnim expresii de genul: "diferența dintre valorile a și b este semnificativă" sau "colectivitatea A diferă semnificativ de B" etc- Trebuie făcută următoarea precizare esențială: Se poate vorbi de semnificația diferenței dintre două valori numai în cazul în care cel puțin una este rezultatul unui studiu pe bază de eșantion. Dacă mărimile respective sînt calculate pe populații întregi, pe care le reprezintă, nu are nici un sens de a vorbi despre semnificația lor (în sensul statistic al acestui termen, evident). Așa de pildă, dacă vîrsta medie a unei colectivități A este de 45 ani, iar a uneia B de 46 ani, diferența de un an este diferența reală și nu are nici un rost să i se aplice calificativul de semnificativă sau nesemnificativă. Dimpotrivă, dacă cele două populații sînt investigate numai selectiv, problema devine reală datorită faptului, cunoscut de noi de pe acum, că prin atari studii se obțin numai evaluări ale mediilor reale, evaluări incluse în intervale de posibilă "fluctuație". Or, este clar că anumite diferențe pot apare numai datorită acestor fluctuații normale oricărei eșantionări. Dimpotrivă, dacă diferențele vor depăși anumite limite, ele nu mai pot fi pu-se pe seama impreciziei eșantionului, ci trebuie să reflecte diferențe reale în populațiile mamă din care eșantioanele provin. Aceste scurte explicații se vor clarifica și preciza în analiza concretă a unor cazuri simple .

A. Primul tip de analiză privește situația cînd una dintre cele două valori care se compară provine dintr-o cercetare selectivă iar cealaltă nu. Să presupunem că este vorba de o medie  $\bar{x}'$  , ce trebuie comparată cu una din populație  $\bar{x}$ . Cum se știe, dacă diferența dintre valori este mai mică decît două erori standard, ea poate fi pusă pe seama fluctuațiilor de eșantionare. Algebric se scrie:

$$\begin{aligned} & |\bar{x} - \bar{x}'| < 2\sigma \\ \text{sau} & \frac{|\bar{x} - \bar{x}'|}{\sigma} < 2 \end{aligned}$$



În general, se calculează mărimea din stînga semnului de inegalitate, notată cu  $z$

$$z = \frac{|\bar{x} - \bar{x}'|}{e}$$

și se testează așa-numita ipoteză nulă: "nu există diferență semnificativă între  $\bar{x}$  și  $\bar{x}'$ ". Vom avea mai multe situații:

- Dacă  $z < 2$ , adică ne aflăm sub pragul de probabilitate de 0,95, ipoteza nulă nu poate fi respinsă, deci nu se poate afirma că există o diferență semnificativă între  $\bar{x}$  și  $\bar{x}'$  (Atenție! Nu faptul că diferența este nesemnificativă, ci că ea nu este semnificativă)

- Dacă  $z > 2$ , înseamnă că ipoteza nulă poate fi respinsă - diferența este semnificativă - cu 95% șanse.

- Dacă  $z$  este mai mare decît valori superioare lui 2, atunci gradul de întemeiere a respingerii ipotezei nule este mai ridicat. De pildă, dacă  $z$  depășește valoare 3.3, respingerea ipotezei nule este aproape o certitudine și despre diferență se spune uneori că este foarte semnificativă, deși calificativul de "foarte" are aici doar rolul de a sugera depășirea unui înalt grad de probabilitate.

Observație importantă. În limbajul statisticii, mai ales cînd e vorba de testele de semnificație, s-a introdus obișnuința ca nivelele de probabilitate utilizate, de exemplu 0,95; 0,99; 0,999 etc. să fie notate prin diferențele acestor probabilități la unitate, adică: 0,05; 0,01; 0,001; etc. Astfel se spune că diferența  $\bar{x} - \bar{x}'$  este semnificativă la nivelul 0,05. Asta înseamnă pur și simplu că sînt 95% șanse ca ea să fie semnificativă.

#### Exemple.

a) Cu ajutorul acestui tip de analiză putem verifica reprezentativitatea eșantionului ales, dacă cunoaștem un parametru din populație. Voi prezenta un caz de acest fel furnizat de lucrarea "Studentul și societatea (1973) scrisă pe baza rezultatelor unei anchete pe un eșantion de 2172 studenți. În cadrul eșantionului a fost găsită o proporție de 54,9% băieți, știindu-se în același timp, că procentul real al studenților de sex masculin pe țară era, la data anchetei, de 55,6%. Întrebarea care se ridică este dacă diferența de 0,7% se poate datora unor fluctuații normale





de eşantionare sau există un factor sistematic care a distorsionat eşantionul. În cazul în care se constată - şi se va vedea imediat că într-adevăr se constată - că diferenţa nu este semnificativă, putem transfera cu destulă siguranţă acest calificativ de corectitudine a eşantionării şi asupra altor caracteristici investigate.

Cu notaţiile noastre vom avea:

$$p = 0,556 \text{ (proporţia în populaţie)}$$

$$p' = 0,549 \text{ (proporţia în eşantion)}$$

$$\sigma_x = \sqrt{0,556(1-0,556)} = 0,497$$

$$e = \frac{0,497}{\sqrt{2172}} = 0,0107$$

$$|p - p'| = 0,556 - 0,549 = 0,007$$

$$z = \frac{|p - p'|}{e} = \frac{0,0070}{0,0107} = 0,65$$

Prin urmare,  $z$  este mai mic decît 2 şi diferenţa nu este semnificativă.

b) Să presupunem că media de vîrstă pentru muncitorii dintr-o ramură industrială este de 31 ani. Pe un eşantion de 625 de muncitori din ramura respectivă, în cadrul unui judeţ A, s-a obţinut o medie de 32 de ani şi o abatere standard de 5 ani. Întrebarea este dacă putem afirma că în judeţul A media de vîrstă este mai ridicată decît pe ansamblul ţării sau diferenţa de 1 an poate fi pusă pe seama unor fluctuaţii normale de eşantionare. Avem deci:

$$\bar{x} = 31; \bar{x}' = 32; \sigma_x = 5$$

Va rezulta:

$$e = \frac{5}{\sqrt{625}} = \frac{5}{25} = 0,2$$

$$z = \frac{|\bar{x} - \bar{x}'|}{e} = \frac{1}{0,2} = \frac{10}{2} = 5$$



Deci  $z$  este mai mare chiar decât 3,3 și putem spune că diferența este semnificativă la nivelul de probabilitate 0,001 (adică sînt 99,9% șanse ca vîrsta medie în județul A să fie mai mare decât pe țară).

B. Al doilea tip de analiză are în vedere compararea valorilor medii (proporțiilor) din două eșantioane diferite (chiar și ca mărime). Avînd două populații  $A_1$  și  $A_2$ , presupunem că am efectuat o cercetare pe cîte un eșantion  $n_1$  (din  $A_1$ ) și  $n_2$  (din  $A_2$ ), obținînd pentru o aceeași variabilă mediile  $\bar{x}'_1$  și  $\bar{x}'_2$ , cu abaterile standard  $\sigma_{x'_1}$  și  $\sigma_{x'_2}$ .

Raționînd ca în cazul unei singure populații, putem și aici imagina o infinitate de cupluri de eșantioane, obținînd pentru fiecare cuplu o variabilă dată de diferența mediilor eșantioanelor.

Teoria statisticii demonstrează că pentru această variabilă eroarea standard este dată de formula:

$$e = \sqrt{\frac{\sigma_{x'_1}^2}{n_1} + \frac{\sigma_{x'_2}^2}{n_2}}$$

Exemplu. Fie  $A_1$  populația studenților de la Universitate și  $A_2$  de la Institutul Agronomic. Să admitem următoarele date obținute prin sondaj:

$$\begin{aligned} n_1 &= 900 & n_2 &= 700 \\ \bar{x}'_1 &= 22 \text{ ani} & \bar{x}'_2 &= 22,5 \text{ ani} \\ \sigma_{x'_1} &= 1,5 \text{ ani} & \sigma_{x'_2} &= 1,6 \text{ ani} \end{aligned}$$

Atunci

$$e = \sqrt{\frac{1,5^2}{900} + \frac{1,6^2}{700}} = 0,08$$

$$z = \frac{|22,5 - 22|}{0,08} = \frac{0,5}{0,08} = \frac{50}{8} = 6,3$$

Valoarea  $z$  depășește și 3,3 deci diferența este semnificativă.



Cu alte cuvinte se poate afirma că la Agronomie media de vîrstă este mai ridicată decît la Universitate.

## 2.2. Testul $\chi^2$

Se folosește mai ales pentru caracteristicile calitative, spre a se stabili semnificația diferențelor între frecvențe. Se pot imagina multe situații de aplicare a testului, dintre care mai frecvente sînt următoarele: a) compararea frecvențelor diferitelor clase ale unei aceleiași caracteristici; b) compararea distribuțiilor a două caracteristici cu același număr de clase; c) stabilirea existenței asocierii într-un tabel de asociere.

În esență, se pornește de la anumite tabele ce cuprind un număr de căsuțe cu frecvențele respective. Să presupunem că avem  $s$  căsuțe cu frecvențele empirice  $n_1, n_2, \dots, n_s$ . Aceste frecvențe se vor compara cu un șir corespunzător de frecvențe numite teoretice (numite așa pentru motivul că adesea ele se deduc dintr-o ipoteză oarecare și nu sînt culese din realitatea empirică):

$m_1, m_2, \dots, m_s$ .

Mărimea  $\chi^2$  e dată de formula:

$$\chi^2 = \frac{(n_1 - m_1)^2}{m_1} + \frac{(n_2 - m_2)^2}{m_2} + \dots + \frac{(n_s - m_s)^2}{m_s}$$

$\chi^2$  este o variabilă statistică pentru care se pot calcula probabilitățile de a obține valori mai mari decît anumite valori date. Mai exact, pentru anumite praguri de probabilitate: 0,05, 0,01 etc. se pot determina valorile  $\chi^2_0$ , valori care nu sînt depășite decît cu probabilitățile respective (ce pot fi oricît de mici). Aceste valori critice, care se găsesc în tabele special întocmite, depind nu numai de probabilități ci și de un alt parametru numit numărul gradelor de libertate al tabelului analizat. Prin numărul gradelor de libertate se înțelege numărul căsuțelor interioare ale unui tabel cu totalurile date, care se pot completa după voie.

Iată un extras din tabelul cu valorile critice ale lui



Nr. grade de libertate( $\nu$ )	$P = 0,05$	$P = 0,01$
1	3,841	6,635
2	5,991	9,210
3	7,815	11,345
4	9,488	13,277
5	11,070	15,088
6	12,592	16,812
7	14,067	18,475
8	15,507	20,090
9	16,919	21,666
10	18,307	23,209

Pentru un grad de libertate, diferența devine semnificativă la nivelul 0,05 dacă  $\chi^2$  depășește pe 3,841 și la nivelul 0,01 dacă depășește 6,635.

Cîteva exemple vor lămuri foarte bine utilizarea testului.

a) Intr-un eșantion de 100 persoane s-au găsit 45 de bărbați și 55 femei. Se poate afirma că în populație sînt mai multe femei decît bărbați ?

Avem deci frecvențele empirice:

$$n_1 = 45; \quad n_2 = 55$$

care se cer comparate cu frecvențele teoretice:

$$m_1 = 50; \quad m_2 = 50$$

obținute în ipoteza coincidenței numerice a indivizilor de sexe diferite. Există un singur grad de libertate, căci dacă fixăm o frecvență, cealaltă rezultă din diferența la total.

$$\chi^2 = \frac{(45-50)^2}{50} + \frac{(55-50)^2}{50} = \frac{25}{50} + \frac{25}{50} = \frac{50}{50} = 1$$

$$\chi^2 = 1 < \chi^2_{0,05}(P = 0,05) = 3,841$$



Deci ipoteza nulă nu poate fi respinsă, adică nu se poate spune că în populație ar fi mai multe femei decât bărbați.

Observație. Trebuie subliniat că variabila  $\chi^2$  face apel la frecvențele absolute, nu la cele relative sau procente. Dacă în cazul de mai sus s-ar multiplica toate cifrele cu 10 (adică dacă eșantionul ar fi fost de 1000, cu 450 bărbați și 550 femei), structura sa ar fi fost aceeași dar valoarea lui  $\chi^2$  ar fi fost 10, în loc de 1, depășind valoarea critică chiar și pentru nivelul 0,01.

b) Conform unor date existente, structura socială a unei colectivități este: 25% agricultori, 50% muncitori, 10% intelectuali și 15% alte categorii sociale, Într-un eșantion de 1000 de persoane s-au obținut frecvențele:

Agricultori	Muncitori	Intelectuali	Alții	Total
230	470	135	165	1000

Este eșantionul bine construit ?

Conform ipotezei, frecvențele teoretice sînt: 250, 500, 100 și 150.

$$\chi^2 = \frac{(230-250)^2}{250} + \frac{(470-500)^2}{500} + \frac{(135-100)^2}{100} + \frac{(165-150)^2}{150}$$

$$= \frac{400}{250} + \frac{900}{500} + \frac{1225}{100} + \frac{225}{150} = 1,6 + 1,8 + 12,25 + 1,5 = 17,15$$

Numărul gradelor de libertate fiind 3, rezultă:

$$\chi^2 = 17,15 > \chi^2_{0,01}(P = 0,01) = 11,345$$

Deci valoarea este semnificativă, ceea ce înseamnă că diferența de structură dintre eșantion și populație nu poate fi pusă pe seama fluctuațiilor de eșantionare, ci pe o procedură distorsionantă.

c) Într-o cercetare se urmărește detectarea unei asocieri între vîrstă și atitudinea față de o anumită problemă. Clasele caracteristicilor și datele ipotetice sînt prezentate în tabelul 1.



Atitudinea Vîrsta	Favorabilă	Neutră	Nefavorabilă	Total
Tineri	100	80	20	200
Adulți	80	100	20	200
Vîrstnici	20	20	60	100
TOTAL	200	200	100	500

Tabelul 1. Frecvențe empirice

Pentru stabilirea unei eventuale asocieri, comparăm tabelul cu frecvențe reale (empirice) cu cel de independență (frecvențele teoretice). Se observă imediat o oarecare concentrare a indivizilor pe diagonala principală. Se ridică întrebarea dacă acest lucru poate sau nu să se datoreze numai fluctuațiilor de eșantionare.

Atitudinea Vîrsta	Favorabilă	Neutră	Nefavorabilă	Total
Tineri	80	80	40	200
Adulți	80	80	40	200
Vîrstnici	40	40	20	100
TOTAL	200	200	100	500

Tabelul 2. Tabelul de independență, construit cu totalurile tabelului 1.

Se calculează:

$$\chi^2 = \frac{(100-80)^2}{80} + \frac{(80-80)^2}{80} + \frac{(20-40)^2}{40} + \frac{(80-80)^2}{80} + \frac{(100-80)^2}{80} + \frac{(20-40)^2}{40} + \frac{(20-40)^2}{40} + \frac{(20-40)^2}{40} + \frac{(60-20)^2}{20}$$



$$= \frac{400}{80} + 0 + \frac{400}{40} + 0 + \frac{400}{80} + \frac{400}{40} + \frac{400}{40} + \frac{400}{40} + \frac{1600}{20} =$$

$$= 5 + 10 + 5 + 10 + 10 + 10 + 80 = 140$$

Numărul gradelor de libertate este 4. La acest număr, mărimea  $\chi^2$  obținută este cu mult mai mare decât valorile critice din tabel, deci asocierea este semnificativă.

Observație . Ne reamintim că mărimea  $\chi^2$  a fost utilizată la calcularea coeficienților de asociere. Este clar că, dacă e vorba de o cercetare selectivă, primul pas ce trebuie făcut este cel de mai sus, adică testarea semnificației asocierii. Abia apoi, cu aceeași mărime  $\chi^2$ , se va trece la calcularea coeficientului de asociere, despre care se va putea spune, la rîndul său, că este semnificativ).





B I B L I O G R A F I E

- ALUAS, I., DRAGAN, I. (1971) Sociologia franceză contemporană.  
Antologie, Ed. Politică, București
- BARBAT, A. I. (1972) Teoria statisticii sociale, Ed. did. și ped.,  
Buc.
- BIJI, M., BIJI, E. (1979) Statistică teoretică, Ed. did. și ped.,  
Buc.
- BLALOCK, H. (1973) Introduction à la recherche sociale, Duculot,  
Gembloux
- BOUDON, R., LAZARSFELD, P. (1965) Le vocabulaire des sciences  
sociales; Concepts et indices, Mouton, Paris
- BOUDON, R., LAZARSFELD, P. (1966) L'Analyse empirique de la causa-  
lité, Mouton, Paris
- BOUDON, R. (1967) L'Analyse mathématique des faits sociaux, Blon,  
Paris
- BOUDON, R. (1971) Les mathématiques en sociologie, P.U.F., Paris
- BOUDON, R. (1973) L'Inégalité des changes, Armand Colin, Paris
- BOURDIEU, P. et al. (1968) Le métier de sociologue, Mouton, Paris
- BOURDIEU, P. et al. (1971) "Abdicățiile empirismului" în I. Aluas  
și I. Drăgan (1971)
- CAPLOW, Th. (1970) L'Enquête sociologique, Armand Colin, Paris
- CHELCEA, S. (1975) Chestionarul în investigația sociologică, Ed.  
științifică și enciclopedică, București
- CHELCEA, S. (1982) Experimentul în psihosociologie, Ed. științifi-  
că și enciclopedică, București
- CHELCEA, S. (coordonator (1985) Semnificația documentelor sociale,  
Ed. științifică și enciclopedică, București
- COSTEA, St. et al. (1983) Sociologia românească contemporană, Ed.  
științifică și enciclopedică, București
- DURKHEIM, E. (1974) Regulile metodei sociologice, Ed. științifică,  
București
- DUVERGER, M. (1964) Méthodes des sciences sociales, P.U.F., Paris
- PESTINGER, L., KATZ, D. (1963) Les méthodes de recherche dans les  
sciences sociales, P.U.F., Paris



- GEORGESCU, P. (1979) Indrumător pentru cercetarea sociologică în cultura fizică, Ed. Sport-Turism, București
- GRAWITZ, M. (1972) Méthodes des sciences sociales, Dalloz, Paris
- HERSENI, T. (1982) Sociologie, Ed. științifică și enciclopedică, București
- LAZARSFELD, P. (1966) L'Interprétation des relations statistiques comme procédure de recherche, în R. Boudon și P. Lazarsfeld (1966)
- LAZARSFELD, P. (1970) Philosophie des sciences sociales, Gallimard, Paris
- LAZARSFELD, P. (1971) Qu'est-ce que la sociologie, Gallimard, Paris
- LOETHER, H., J. McTAVISH, D. G. (1976) Descriptive and Inferential Statistics, Allyn and Bacon, Boston
- MARGINEAN, I. (1982) Măsurarea în sociologie, Ed. științifică și enciclopedică, Buc.
- MERTON, R. K. (1972) Influența cercetării empirice asupra teoriei sociologice în Cunoașterea faptului social, seria Teorie și metodă și științele sociale, vol. VIII, Ed. politică, București
- MIPTODE, V. (1982) Introducere în metodologia investigației sociologice, Ed. Junimea, Iași
- MIHOC, Gh. et al (1982) Modele de analiză statistică, Ed. științifică și enciclopedică, București
- MIHU, A. (1973) ABC-ul investigației sociologice, vol. III, Ed. Dacia, Cluj
- MOSER, C. A. (1967) Metode de anchetă în investigarea fenomenelor sociale, Ed. științifică, București
- NICOLA, T. (1973) Experimentul social, Ed. Scrisul românesc, Craiova
- NOELLE, E. (1966) Les sondages d'opinion, Minuit, Paris
- PAGES, R. (1971) Experimentul în sociologie, în I. Aluș, I. Drăgan (1971)
- ROTARIU, T. (1974) Influența factorilor mediu de proveniență, origine socială și pregătire școlară asupra reușitei la admiterea în învățământul superior, în Studia, Series Sociologia.
- SHILS, E. (1967) Tendințele cercetării sociologice, în vol. Sociologia contemporană



SOROKIN, P.A. (1967) Diversitate și unitate în sociologie, în vol.

Sociologia contemporană

STAHL, H.H. (1974) Teoria și practica investigațiilor sociologice,  
vol. I. și II, Ed. științifică și enciclopedică, Buc.

STOETZEL, J., GIRARD, A. (1975) Sondajele de opinie publică, Ed.  
științifică și enciclopedică, București

SZCZEPANSKI, J. (1972) Noțiuni elementare de sociologie, Ed. științifică,  
Buc.

VLASCEANU, L. (1982) Metodologia cercetării sociologice, Ed. științifică  
și enciclopedică, București

YEOMANS, K.A. (1968) Statistics for the social scientist, vol. I și  
II, Penguin Books, Harmondsworth

YULE, G.U., KENDAKK, M.G. (1969) Introducere în teoria statisticii,  
Ed. științifică, București

ZAMFIR, C. et al. (1980) Dezvoltarea umană a întreprinderii, Ed.  
Academiei RSR, București

x x x (1967) Sociologia contemporană, în seria Teorie și metodă  
în științele sociale, vol. V., Ed. politică, București

x x x (1973) Studentul și societatea, Ed. Dacia, Cluj